TKK Dissertations 135
Espoo 2008

NOVEL GENETIC FITTING ALGORITHMS AND STATISTICAL ERROR ANALYSIS METHODS FOR X-RAY REFLECTIVITY ANALYSIS

Doctoral Dissertation

Jouni Tiilikainen

Helsinki University of Technology
Faculty of Electronics, Communications and Automation
Department of Micro and Nanosciences

TKK Dissertations 135
Espoo 2008

NOVEL GENETIC FITTING ALGORITHMS AND STATISTICAL ERROR ANALYSIS METHODS FOR X-RAY REFLECTIVITY ANALYSIS

Doctoral Dissertation

Jouni Tiilikainen

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Faculty of Electronics, Communications and Automation for public examination and debate in Auditorium AS1 at Helsinki University of Technology (Espoo, Finland) on the 3rd of October, 2008, at 12 noon.

Helsinki University of Technology
Faculty of Electronics, Communications and Automation
Department of Micro and Nanosciences

Teknillinen korkeakoulu
Elektroniikan, tietoliikenteen ja automaation tiedekunta
Mikro- ja nanotekniikan laitos

Distribution:
Helsinki University of Technology
Faculty of Electronics, Communications and Automation
Department of Micro and Nanosciences
P.O. Box 3500 (Tietotie 3)
FI - 02015 TKK
FINLAND
URL: http://www.micronova.fi/units/mns/
Tel. +358-9-4511
Fax +358-9-451 3128
E-mail: jouni.tiilikainen@tkk.fi

ISBN 978-951-22-9540-1
ISBN 978-951-22-9541-8 (PDF)
ISSN 1795-2239
ISSN 1795-4584 (PDF)
URL: http://lib.tkk.fi/Diss/2008/isbn9789512295418/

TKK-DISS-2503

Suomen Graafiset Palvelut Oy Ltd
Kuopio 2008

HELSINKI UNIVERSITY OF TECHNOLOGY
P. O. BOX 1000, FI-02015 TKK
http://www.tkk.fi

ABSTRACT OF DOCTORAL DISSERTATION

Author Jouni Tiilikainen

Name of the dissertation

Novel genetic fitting algorithms and statistical error analysis methods for x–ray reflectivity analysis

Date of manuscript	May 10th, 2008	Date of the dissertation	October 3rd, 2008
Article dissertation (summary + original articles)		Number of pages	65 + 37
Faculty	Faculty of Electronics, Communications and Automation		
Department	Department of Micro and Nanosciences		
Field of research	Nanotechnology		
Opponent	Dr. Joachim Woitok		
Supervisor	Professor Harri Lipsanen		

Abstract

This thesis deals with x–ray reflectivity (XRR) analysis. XRR is a very accurate technique for the metrology of thin films but the analysis of measurements has been difficult thus limiting every day material research. In this thesis, novel genetic algorithms (GAs) for XRR curve fitting and statistical error analysis methods are developed. The XRR analysis utilizes very accurate Parratt's formalism combined with Nevot–Croce interface roughness. The analysis concentrates on the atomic layer deposited materials by using models mimicking their properties. The properties of GAs are studied using aluminium oxide/zinc oxide nanolaminate models. Models of aluminium oxide layers on silicon substrate are used in the case of the error analysis.

The demonstrated novel GAs are utilizing the rotation of coordinates during the crossover phase to reduce interparameter dependencies. The new basis is formed from the eigenvectors of Hessian and statistical covariance matrices. The crossover is performed in the rotated coordinates and the new combinations are transformed back to the original coordinates. It is shown that the coordinate rotation improves the convergence properties of GAs in complex XRR curve fitting problems and a statistical approach is more powerful than the Hessian matrix method. Furthermore, a GA using independent component analysis gives additional robustness to the curve fitting by utilizing a nonorthogonal linear transformation technique.

The interdependency of XRR parameters is studied using fitness landscapes. The fitness landscape analysis utilizes subspace projection of the original parameter space where the projection is done using an experimental model. The work reveals that the error in the determined mass density can compensate the error in surface roughness thus diminishing the accuracy of both of these parameters. This result is also verified later with other methods.

The effect of Poisson noise on the accuracy of XRR analysis is studied statistically. Thickness determination accuracy of an aluminium oxide layer is ± 0.09 nm with 99% confidence in the studied case which represents the lower limit for the error. Here the analysis assumed a perfect fit to the measurement. The upper error is achieved by taking into account a nonideal fit by separating the effect of noise from the fitness value. In a case of the studied measurement, the determined thickness error is ± 0.12 nm with 99% confidence.

Keywords Atomic layer deposition, Genetic algorithm, Fitting, Error analysis, X–ray reflectivity

ISBN (printed)	978–951–22–9540–1	ISSN (printed)	1795–2239
ISBN (electronic)	978–951–22–9541–8	ISSN (electronic)	1795–4584
Publisher	TKK Department of Micro and Nanosciences		
Print distribution	TKK Department of Micro and Nanosciences		

The dissertation can be read at http://lib.tkk.fi/Diss/2008/isbn9789512295418

TEKNILLINEN KORKEAKOULU PL 1000, 02015 TKK http://www.tkk.fi	VÄITÖSKIRJAN TIIVISTELMÄ

Tekijä Jouni Tiilikainen

Väitöskirjan nimi

Novel genetic fitting algorithms and statistical error analysis methods for x–ray reflectivity analysis

Käsikirjoituksen jättämispäivämäärä 10.5.2008	Väitöstilaisuuden ajankohta 3.10.2008
Yhdistelmäväitöskirja (yhteenveto + erillisartikkelit)	Sivumäärä 65 + 37

Tiedekunta	Elektroniikan, tietoliikenteen ja automaatiotekniikan tiedekunta
Laitos	Mikro– ja nanotekniikan laitos
Tutkimusala	Nanoteknologia
Vastaväittäjä	Dr. Joachim Woitok
Työn valvoja	Prof. Harri Lipsanen

Tiivistelmä

Väitöskirja käsittelee röntgenheijastusanalyysia. Röntgenheijastus on hyvin tarkka menetelmä ohutkalvojen mittauksissa mutta mittausten analyysi on ollut vaikeaa rajoittaen materiaalitutkimusta. Materiaalitutkimuksen työkaluiksi tässä työssä kehitetään geneettisiä sovitusalgoritmeja ja virheanalyysimenetelmiä. Analyysissa hyödynnetään tarkkaa Parratt'n menetelmää Nevotin ja Crocen karheusapproksimaatiolla. Väitöstyössä käytetään atomikerroskasvatettuja kerrosrakenteita malleina sovituskäyrien laskennassa. Sovitusalgoritmien ominaisuuksia tutkitaan käyttäen alumiini-/sinkkioksidi–nanolaminaatteja malleina ja virheanalyysissä käytetään yksinkertaisia alumiinioksidirakenteita piin päällä.

Työssä esitellään uusia geneettisiä sovitusalgoritmeja, jotka käyttävät koordinaatiston rotaatiota risteytymisen aikana, millä vähennetään parametrien välisiä riippuvuuksia. Rotaatiokoordinaatisto saadaan Hessianin ja tilastollisen kovarianssimatriisin ominaisvektoreista. Risteytyminen suoritetaan uudessa koordinaatistossa ja risteymä muunnetaan takaisin alkuperäiseen koordinaatistoon. Työssä näytetään, että koordinaatiston rotaatio parantaa geneettisten algoritmien suppenemisominaisuuksia ja tässä tapauksessa tilastollinen lähestymistapa on parempi. Lisäksi työssä osoitetaan, että riippumattomien komponenttien analyysia hyödyntävä geneettinen algoritmi parantaa entisestään sovituksen onnistuvuutta.

Sovitettavien parametrien ristivaikutuksia tutkitaan käyttäen hyvyyskarttoja, jotka kuvaavat sovituksen hyvyyttä käytettyjen parametrien funktiona. Työssä hyödynnetään kokeellista mallia, jolla projisoidaan ongelma aliavaruuteen ja näin helpotetaan analyysia. Työssä osoitetaan, että massatiheyden määrityksessä tehty virhe voidaan kompensoida pinnan karheuden virheellä, mikä vähentää röntgenheijastuksen tarkkuutta.

Poissonin kohinan vaikutusta sovituksen hyvyyteen tutkitaan tilastollisesti olettamalla täydellinen sovitus mittaukseen. Paksuuden virheeksi määritettiin $\pm 0{,}09$ nm tutkitulle tapaukselle 99% luottamustasolla, mikä edustaa virheen alarajaa. Virheen yläraja saadaan ottamalla huomioon sovituksen epäideaalisuus erillistämällä hyvyyslukuun vaikuttava kohina. Tässä tapauksessa tutkittiin aitoa mittausta ja virherajoiksi saatiin $\pm 0{,}12$ nm 99% luottamustasolla.

Asiasanat Atomikerroskasvatus, geneettiset algoritmit, sovittaminen, virheanalyysi, röntgenheijastus

ISBN (painettu)	978–951–22–9540–1	ISSN (painettu)	1795–2239
ISBN (elektroninen)	978–951–22–9541–8	ISSN (elektroninen)	1795–4584
Julkaisija	TKK Mikro– ja nanotekniikan laitos		

Painetun väitöskirjan jakelu TKK Mikro– ja nanotekniikan laitos

Luettavissa verkossa osoitteessa http://lib.tkk.fi/Diss/2008/isbn9789512295418

Preface

This work has been done in the Department of Micro and Nanosciences (former Micro and Nanosciences Laboratory) in Helsinki University of Technology during 2005–2008. I wish to express my gratitude to Professor Harri Lipsanen for the opportunity to work in his Nanotechnology Group.

I would like to thank my collaborators, Docent Teppo Hakkarainen, Dr. Marco Mattila, Dr. Jaakko Sormunen, Markus Bosund and Juha–Matti Tilli for their patience to try and discuss about my ideas in the field where none of us have proper background education. I am grateful to Professor Seppo Honkanen, Docent Markku Sopanen, Dr. Veli–Matti Airaksinen, Pasi Kostamo and Pekka Törmä for the valuable support in miscellaneous points of my path. In addition, I want to thank also the peer students of the old Optoelectronics Laboratory for the encouragement and good spirit.

For the personal financial support I acknowledge The Graduate School of Electrical Engineering and Tekniikan edistämissäätiö. I also acknowledge Academy of Finland and Finnish Agency for Technology and Innovation (TEKES, ALDUS Project) for the general financial support which made it possible to carry out this thesis. The Finnish IT Center for Science (CSC) is acknowledged for providing computational resources. I thank also Beneq for the fruitful collaboration.

Finally, I would to thank my parents and my friends for their support in my private life. Especially I am grateful to my girlfriend Saila for her patience with my endless and interesting (not!) considerations during these years.

Espoo, February 2008

Jouni Tiilikainen

Contents

List of publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I J. Tiilikainen, J.–M. Tilli, V. Bosund, M. Mattila, T. Hakkarainen, V.–M. Airaksinen and H. Lipsanen, *Nonlinear fitness–space–structure adaptation and principal component analysis in genetic algorithms: an application to x–ray reflectivity analysis*, Journal of Physics D: Applied Physics **40** (2007) 215–218

II J. Tiilikainen, V. Bosund, J.–M. Tilli, J. Sormunen, M. Mattila, T. Hakkarainen and H. Lipsanen, *Genetic algorithm using independent component analysis in x–ray reflectivity curve fitting of periodic layer structures*, Journal of Physics D: Applied Physics **40** (2007) 6000–6004

III J. Tiilikainen, V. Bosund, M. Mattila, T. Hakkarainen, J. Sormunen and H. Lipsanen, *Fitness function and nonunique solutions in x–ray reflectivity curve fitting: Crosserror between surface roughness and mass density*, Journal of Physics D: Applied Physics **40** (2007) 4259–4263

IV J. Tiilikainen, J.–M. Tilli, V. Bosund, M. Mattila, T. Hakkarainen, J. Sormunen and H. Lipsanen, *Accuracy in x–ray reflectivity analysis*, Journal of Physics D: Applied Physics **40** (2007) 7497–7501

V J. Tiilikainen, M. Mattila, T. Hakkarainen and H. Lipsanen, *Novel method for error limit determination in x–ray reflectivity analysis*, Journal of Physics D: Applied Physics **41** (2008) 115302

Author's contribution

The author has written the manuscripts for all the publications. He designed and realized all the work except the experiments for publication III which were performed by J.–M. Tilli and V. Bosund with assistance of the author. For publications I–IV J.–M. Tilli and for all publications M. Mattila contributed to improve the theoretical basis of the manuscripts together with the author. V. Bosund contributed as the skilled experimentalist for publications I–IV. For publications II, III and IV, J. Sormunen and for all publications T. Hakkarainen were contributing as the proof readers of the manuscripts. For publication I, V.–M. Airaksinen and for all publications H. Lipsanen were supervisors.

List of abbreviations

ALD	Atomic layer deposition
AlO	Aluminium oxide
CGA	Classical (or conventional) genetic algorithm
CovGA	Genetic algorithm using covariance matrix
DEZn	Diethylzinc
EFICA	Efficient FastICA
FastICA	Algorithm for fast independent component analysis
GA	Genetic algorithm
HGA	Genetic algorithm using Hessian matrix
ICA	Independent component analysis
ICAGA	Genetic algorithm using independent component analysis
NL–FSSA	Nonlinear fitness–space–structure adaptation
PCA	Principal component analysis
PCAGA	Genetic algorithm using principal component analysis
TMA	Trimethylaluminium
XGA	eXtended genetic algorithm
XRR	X–ray reflectivity
ZnO	Zinc oxide

1 Introduction

Atomic layer deposition (ALD) [1, 2], a conformal thin film coating method, has reached wide attention in IC industry and on nanoscience field in general due to its capability to fabricate conformal pinhole–free layers with the thickness control in subnanometer range [3]. The conventional application of ALD has been in the fabrication of materials for electroluminescence displays for last decades but nowadays ALD has emerged for 45 nm CMOS technology [4–6]. Furthermore, ALD has obtained wider attention due to its flexibility for a wide class of applications, for instance, in the fabrication of visually invisible protective layers to prevent tarnishing, insulators [7] and conductive [8] materials for transparent electronics, hard coatings [9], diffusion barriers [10], materials compatible with plastics [11, 12], synthetic functional nanolaminates [13] and many others [14, 15].

Thickness determination is a challenge in the field of thin film coating. The oldest approach for thickness determination of ALD samples is to check an interference colour visually. The approach is especially useful when the uniformity of layers should be detected from large substrate areas but the method is imprecise in the determination of absolute thickness. One limiting factor in the accuracy is the altering interference colour as a function of refractive index but one can somewhat circumvent this problem by using ellipsometry. However, the refractive index can also change as a function of thickness [16–18] or depth [19] and therefore a separate determination of the optical properties of film material is still necessary for accurate thickness determination in some cases [20]. This so–called thickness–optical property correlation can be overcome by using x–ray reflectivity (XRR [21, 22]) for the thickness determination and subsequently, the fixed XRR thickness in the ellipsometric analysis [23]. Afterwards ellipsometry can, in principle, be used for materials having similar optical properties for accurate thickness determination. Although this approach has still some correspondence problems between ellipsometry and XRR results [24–26], it shows that XRR has an important role in thickness measurements.

The major benefit of XRR is very good accuracy in the thickness determination due to short wavelength of x–rays. In addition to the thickness, the method also provides information from the mass density and interfacial roughnesses of the layer structure. The analysis of reflectivity measurements is not simple and a plethora of analysis methods has been developed. The techniques include phaseless inverse scattering

methods [27–29], Fourier analysis related techniques [30–33], the moments analysis method [34], groove tracking method [35–42], wavelet analysis [43–45], the method based on distorted Born approximation [46], the successive descent method [47], the maximum entropy method [48], the Bayesian approach for reflectivity data analysis [49], techniques using parametric B–splines [50] or Chebyshev series [51] and the technique mixing different methods [52]. Some of these methods are reviewed in [53].

The above mentioned techniques are, however, less accurate than the analysis based on Parratt's formalism [54] with roughness approximations [55]. In this approach, a theoretical curve is fitted to the measurement using techniques such as simulated annealing [51, 56–58], the simplex method [50, 56] or evolutionary methods [56, 59–63]. It is worth mentioning that these latter techniques are protected by several patents [64–68]. Despite the recent development in accurate XRR analysis, it is not uncommon that scientists fail to fit even the simplest XRR curves to a measurement in every day research. A lot of time is wasted into the trial–and–error fitting procedure which could be automatized to be fast and repeatable. Also rarely studied ambiguity in XRR [58, 69] has been a significant topic to be investigated. The accuracy of XRR is known by skilled experimentalists on an intuitive level but a precise formulation or a method for objective error limit determination has been missing. Since the potential of XRR has not been fully exploited, the analysis of XRR measurements has been restricted thus limiting material research.

In this thesis, the fitting of XRR curves and the error analysis is developed using ALD layer models in XRR curve calculations. The main focus of this thesis was to improve the performance of fitting algorithms for XRR analysis in the case of very complex layer structures. The other goal was to develop objective methods for XRR error analysis using simple layer structures as test cases. The thesis consists of five chapters after this introduction. Chapter 2 discusses atomic–layer–deposited materials and XRR. In chapter 3, the background for the novel fitting algorithms and the novel error analysis methods is presented. The results based on the background are presented in section 4. Finally, a summary is given in chapter 5. The publications included in this dissertation can be found after the references.

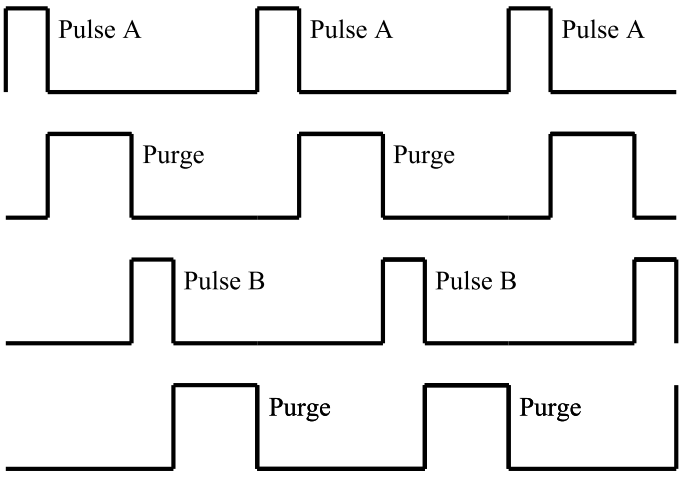


Figure 2.1: Sequential gas phases in ALD cycles. Pulse A means the introduction of gas A. Purge means the evacuation of precursors. Pulse B introduces the other material component to the reactor and subsequent purge removes unreacted species.

2 Experimental background

This chapter introduces the atomic–layer–deposited materials and the characterization technique utilized in this work. Section 2.1 outlines studied ALD materials and section 2.2 introduces the details of XRR.

2.1 Atomic layer deposition and materials

In atomic layer deposition the source gases are introduced sequentially [70] and the growth is digitally controlled, as illustrated in figure 2.1. The simplest models describing the growth of ALD films assume that precursors have infinite reactivity and therefore gas distribution and temperature are not affecting the layer growth. However, the exact growth rate depends, *e.g.*, on the growth temperature, concentration of the precursors in the reactor, surface chemistry in the beginning of the growth, precursor etching and on hardware dependent factors [70–72].

Nonideal growth can degrade the quality of the film by affecting the local layer growth or can change the growth per cycle. Hardware can, for instance, affect the

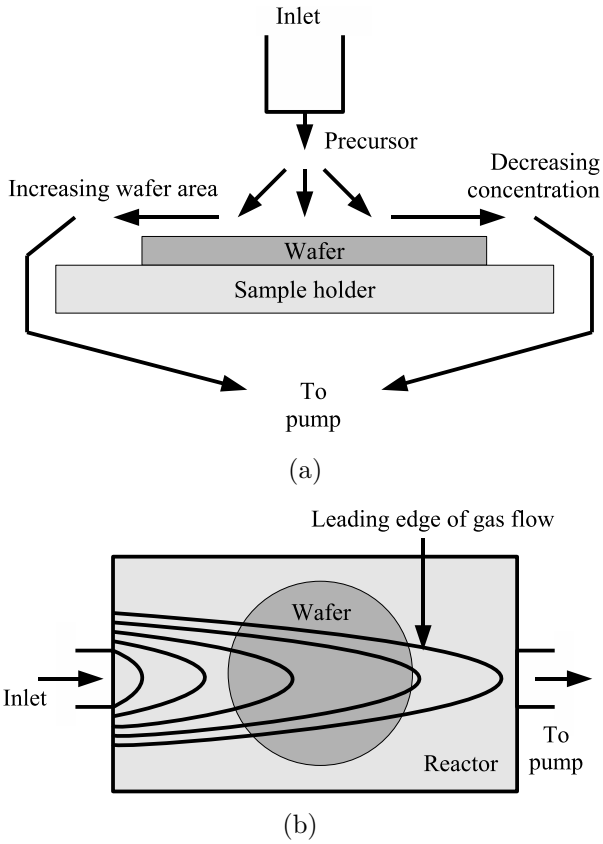


Figure 2.2: Schematic of gas flows in (a) perpendicular and (b) crossflow reactors.

local growth rate via a nonuniform heating profile. In principle, the sample should be as small as possible compared to the dimensions of the reactor, otherwise heat gradients from the heaters to heat sinks affect the precursor reactivity. Possible heat sinks are, *e.g.*, metallic holders connected to the cold walled rack.

Another significant factor causing nonideal growth in local scale is nonuniformly distributed source gases. In the perpendicular type design, shown in figure 2.2(a), the precursors are introduced above the center of the wafer. Although the geometry is beneficial due to its simplicity, the disadvantage in the deposition is the decreasing concentration along the radius of the wafer meanwhile the unit area needing to be coated in the radial direction increases. In a crossover reactor, shown in figure 2.2(b), the precursors are introduced parallel to the wafer surface. The difficulty with this design is to get laminar flow of gases without abrupt edges or gas pockets causing intermixing of precursors in gas phase. In practice this is difficult since the flow profile is nonuniform and some special designs are required to obtain nearly uniform gas distribution in the reactor. Naturally these designs are not perfect which is seen as local variations in the thickness.

One of the factors causing large scale nonlinear growth of thickness is choking, for

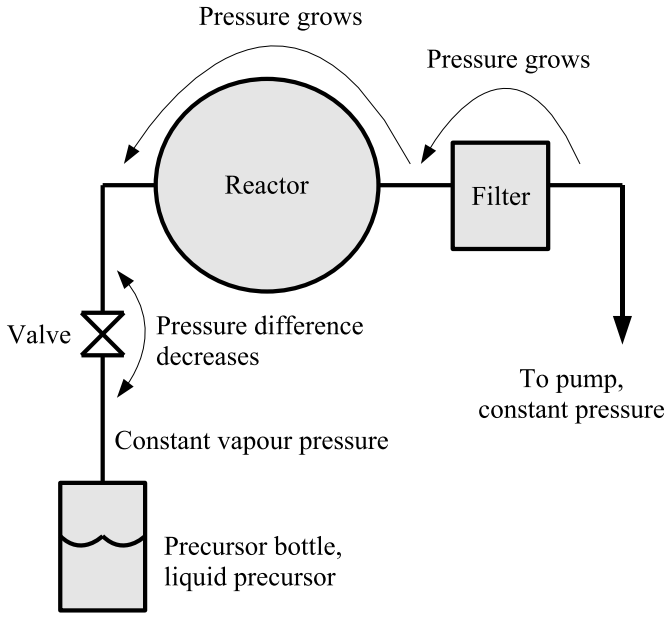


Figure 2.3: Choking of a filter in ALD. Waste in the filter increases the pressure troughout the system and decreases the pressure difference between the lines and the vapour pressure of the liquid precursor.

instance. Figure 2.3 demonstrates the increase of pressure in ALD due to choking of a filter. Overdosing of precursors, which is necessary in ALD fabrication, increases dirt in the filter and the pressure in gas lines drifts up. Since the partial pressure of the evaporated liquid precursor depends on the pressure difference between the vapour pressure of the liquid and the line pressure, the concentration of the precursor in each pulse drifts down with time. Similar drifting is caused by the decreasing partial pressure of the vapour precursor if the time between pulses is not enough for the precursor to reach equilibrium between liquid and gaseous forms. These factors cause the layer growth to be slightly sublinear. The regular maintenance of hardware plays a role if repeatable growth per cycle is kept constant. Maintenance affects, however, the exact pressure drop in the filter and verification measurements should be done to find a new operational point for the hardware. Thus accurate thickness determination methods as verification tools are needed for the precise determination of growth per cycle. Although the growth per cycle may vary along the wafer, the variation is less detrimental with ALD than with other fabrication methods, like in chemical vapour deposition methods. Other significant benefit of ALD is almost ideally growing oxide materials which are not vulnerable for the oxidation of surfaces in atmosphere. Thus interfacial layers are not causing systematic errors in determined thicknesses.

As the above discussion showed, one can avoid the most of the growth nonidealities by a careful fabrication of an ALD film. In such a case, an uniform and pinhole–free layer with the accurate thickness control is obtained and provided for the study

of characterization methods. In this thesis two kinds of ideal ALD materials were studied using x–ray reflectivity. One was aluminium oxide (AlO) deposited on silicon substrates in trimethylaluminium (TMA)/water process. This process is well known both on a theoretical and practical level [2]. The theoretical understanding of the structural properties of AlO were extensively utilized in publications III–V where ambiguity of solutions was studied from different viewpoints. Note that AlO has approximately composition of Al_2O_3 but the expression AlO is used to separate amorphous aluminium oxide from sapphire for clarity.

The other material studied in this thesis was ZnO fabricated using diethylzinc (DEZn)/water process [73]. ZnO together with AlO can be used in the fabrication of ZnO/AlO nanolaminates where the structural properties can be used to tune the conductivity [74]. The structural properties can be tuned by changing the ratio of ZnO and AlO cycles and the growth temperature thus allowing a wide range of tunability [74–76]. However, the dependence of conductivity on these parameters is very complex [74] since interfacial effects, such as nonideal growth [74, 75] and the formation of a nucleation layer [75, 76] are taking place. The structural properties of nanolaminates in general can be studied *ex situ* using XRR [33, 77–84]. However, the fitting of a theoretical curve to the measured curve can be very difficult. In this thesis the efficiency of the fitting procedure with novel genetic algorithms was studied using ZnO/AlO layer models as test cases.

2.2 X-ray reflectivity

2.2.1 Measurement

X–ray reflectivity measurements in this thesis were carried out with a Philips X'Pert Pro instrument shown in figure 2.4. The setup follows a geometry presented in figure 2.5 where the final measurement is carried out using angles $\omega = \theta$. The geometry requires an alignment procedure which contains typically the following steps:

- X–ray beam is measured with a detector in the $\omega = \theta$ geometry without a sample with direct beam. This scan is performed in this geometry to remove an instrumental offset near zero angle.
- The sample on the holder plane is moved in z–direction to set the incident beam to illuminate the sample.
- ω scan is performed to align the sample surface to be parallel with the beam.
- Ψ scan is performed to correct possible misalignment of the sample.

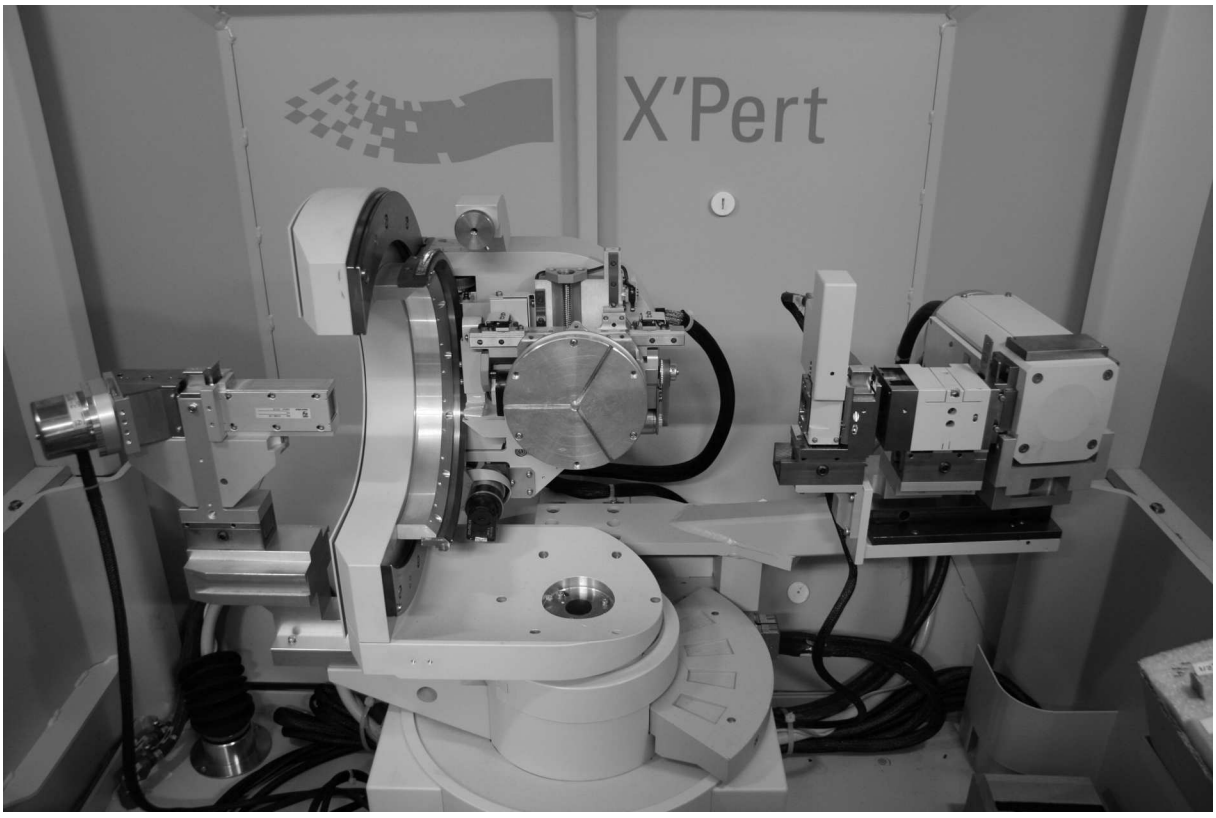


Figure 2.4: Hardware setup used in x–ray reflectivity measurements.

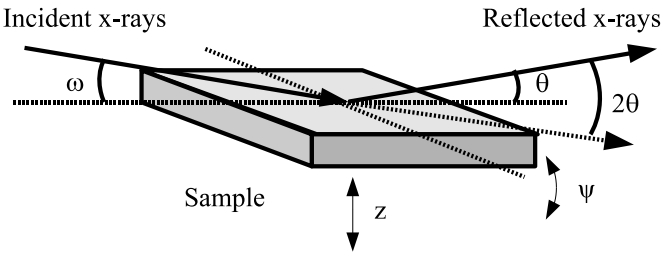


Figure 2.5: Geometry used in XRR measurements.

This procedure shown above should be performed properly since it is critical in the mass density determination of a material [85]. However, the exact geometry varies between instruments and instructions for the alignment procedure are thus apparatus dependent. After the alignment procedure, the angles of incident and reflected beams are kept equal in XRR measurements. XRR measurements require caution and at least following sources for nonidealities need attention:

Stability of the beam

The measured photon count typically varies in the range of several orders of magnitude. Therefore the modulation of the power of the x–ray source may be an attractive way to decrease the number of photons in high reflection regions and increase them in the low reflection regions, respectively. However, the number of emitted photons can be unstable as a function of power and the power setting for the x–ray source should be kept fixed.

In this work the power setting used in the measurements were 40 kV for voltage and 40 mA for current.

Monochromaticity

Monochromaticity of x–rays must be ensured since analysis methods typically assume a single wavelength. Typically this condition can be ensured by using a monochromator.

In this work the CuKα radiation and a flat crystal monochromator were used in the measurements. The manufacturer gives a rocking curve width of 0.4° for the crystal used in the monochromator.

Angular resolution of the goniometer

XRR measurement from a thick layer forms rapidly oscillating reflection pattern as the function of the incident beam angle. Thus the angular resolution of a goniometer affects the maximum thickness which can be measured.

In this thesis the angular resolution of 1/1000° to 1/100° was used in the measurements.

Divergence of the beam

Divergence of the beam affects the instrumental convolution and it is difficult to model accurately with standard XRR analysis methods. The divergence can be reduced by using x–ray mirrors without losing the intensity of beam, for instance, but slits collimating the beam reduce intensity remarkably.

In this work a 1/32° (and sometimes a 1/16°) divergence slit was used in the incident beam setup. The divergence of the reflected beam was reduced by a thin film

collimator and a 0.04 radian soller slit.

Illuminated area

The area illuminated by the incident beam varies as a function of the incident beam angle. The area is nearly infinitely large for the zero angle but decreases rapidly with the increasing angle. The illumination area, in principle, should be reduced to be as small as possible with the incident beam optics. This requirement is due to the analysis methods typically assuming a uniform layer structure in the lateral direction. Therefore the smaller beam divergence improves the correspondence between the measurement and theory. The trade–off here is that the limited area is obtained with the expense of decreased photon counts which increases noise in the measurement and the inaccuracy of the XRR analysis. In practice one should make a compromise between the number of photon counts and the reduction of the illumination area. The illumination area can be somewhat increased by increasing the beam divergence, if the layer is very uniform in lateral direction. This is in fact the case with ALD materials, especially for aluminium oxide and zinc oxide studied in this thesis.

Dynamical range of the detector

The total reflection of x–rays is located at the high intensity region and contains material specific information. One can use, for instance, the angular shift of the half maximum intensity to observe mass density variations between samples having similar stoichiometry. This high intensity region, however, can saturate a scintillation detector and therefore appropriate attenuators should be used to reduce the intensity of the beam down to a reasonable level. The improperly determined attenuation factor and a too high intensity level can shape the high intensity region of the curve and therefore the measurements must be carried out with care to avoid saturations.

The measurements carried out in this thesis were done using the scintillation detector which had a linear response ranging from zero to 400,000 detected photon counts. In the high intensity region, the linear intensity response was achieved by using a 0.1 mm thick copper attenuator in the incident beam optics.

2.2.2 Poisson noise

Noise decreases the accuracy of XRR and it is dependent on principal parameters used in measurements. The parameters are the number of points in the angle range

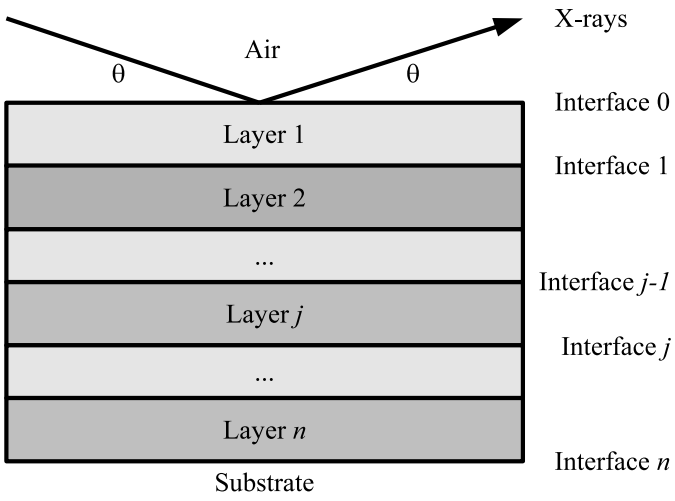


Figure 2.6: Schematic layer structure and the angles of incident and measured x–rays.

and the averaging time per a measured point. The averaging time affects the number of detected photons in one point following the Poisson distribution:

$$P(x_i) = \frac{\overline{x}_i^{x_i}}{x_i!} e^{-\overline{x}_i}, \tag{2.1}$$

where $P(x_i)$ is the probability distribution of the detected counts x_i and $\overline{x}_i$ is the expected number of photon counts for the ith point. The expected number of counts $\overline{x}_i = \overline{I}T$, where T is the averaging time and $\overline{I}$ is the expected intensity which can be calculated using the formalism presented in the following subsection.

2.2.3 Parratt's formalism

The XRR setup measures a layer structure using the geometry where the measured and the incident angles of x–rays are the same. The angles of the geometry and the layer structure required for understanding of the subsequent presentation of the Parratt's formalism [54] are introduced in figure 2.6.

Here the formalism is reviewed in the modified form for the structure containing $j = 0 \ldots n$ layers, where $j = 0$ represents the 'Air layer'. The recurrent reflection coefficient for the amplitude of the electric field of the $(j-1)$th layer

$$r_{j-1} = \frac{F_{j-1} + r_j e^{-\mathrm{i}2k_j t_j}}{1 + F_{j-1} r_j e^{-\mathrm{i}2k_j t_j}}, \tag{2.2}$$

where t_j is the layer thickness and the Fresnel factor

$$F_{j-1} = \frac{k_{j-1} - k_j}{k_{j-1} + k_j}. \tag{2.3}$$

Here

$$k_j = \frac{2\pi}{\lambda}\sqrt{\theta^2 - 2\delta_j - 2\mathrm{i}\beta_j} \tag{2.4}$$

where λ is the wavelength of x–rays (in this thesis the wavelength of CuKα_1 emission, 0.154056 nm, was used) and θ is the angle of the incident and reflected beams. Additionally

$$\delta_j = \frac{N_a r_e \lambda^2}{2\pi}\frac{Z_j + \Delta f_j'}{A_j}\rho_j \tag{2.5}$$

and

$$\beta_j = \frac{N_a r_e \lambda^2}{2\pi}\frac{\Delta f_j''}{A_j}\rho_j, \tag{2.6}$$

where r_e is the Bohr radius, N_a the Avogadro number, Z_j is the average atomic number for the jth layer, A_j is the average atomic mass for the jth layer, $\Delta f_j'$ and $\Delta f_j''$ are material specific dispersion correction coefficients and ρ_j is the mass density of the layer j.

The recurrent computation of the intensity reflection coefficient $R_0 = |r_0|^2$ is begun from the interface n separating the bottom layer and the substrate. For the interface $j = n$, one sets $r_n = F_n$ and continues the calculation of the r_{j-1}'s using equation 2.2. Interfacial roughness can be taken into account using the Nevot–Croce approximation [86]. Here the Fresnel factor is replaced with a term

$$\widetilde{F}_j = F_j e^{-2k_j k_{j+1}\sigma_j^2}, \tag{2.7}$$

where σ_j is the root mean square of the interface roughness amplitude (abbreviated as roughness in this work for the simplicity) of jth interface. Other approximations for the roughness [55] exist but those are less utilized. Using the above mentioned formalism and a proper layer model describing a sample, one can get an agreement between a theoretical curve and the measurement with the resolution down to 0.1 nm [21].

3 Principles of the novel x–ray reflectivity analysis

This chapter discusses the principles of the novel XRR curve fitting methods and the novel error analysis. Section 3.1.1 discusses curve fitting and genetic algorithm. Principles of the coordinate rotation technique and independent component analysis are given in section 3.1.2 and in 3.1.3, respectively. Section 3.2.1 discusses the properties of fitness function used in XRR analysis. Section 3.2.2 introduces a concept called fitness landscape for fitness function analysis. In section 3.2.3, the effect of Poisson noise on the accuracy of fitness is discussed. Finally, section 3.2.4 introduces the principle of the error limit determination in XRR.

3.1 Novel x–ray reflectivity curve fitting methods

3.1.1 Curve fitting and a genetic algorithm

In XRR analysis, a theoretical curve calculated by Parratt's formalism has to be fitted to the measurement. One of the most utilized techniques for fitting is the Levenberg–Marquadt method [87] but its major disadvantage is high dependence on an initialization which can lead the algorithm to converge to a local optimum instead of a global optimum. The convergence to the local optimum can be avoided by using stochastic techniques instead. One of the most popular stochastic global search methods is simulated annealing [88] which mimics in the fitting process the behaviour of atoms in a lattice during annealing. Its main advantage is the guaranteed convergence in a global optimization [89, 90]. In practice, the simulated annealing may require unnecessarily long computation time [91].

A parallel method, called genetic algorithm (GA) [92], can reduce the probability of an algorithm to use time for the search of a false solution, and thus reduce computation time significantly [56]. Figure 3.1 shows the schematics of GA. The method is based on evolution mimicking Darwinian survival–of–the–fittest principle. The principle means that new individuals are formed utilizing the principle inspired by the crossover mechanism between two parents' DNA chromosomes, *i.e.*, a part of genes is copied from the first parent while the rest of the genes are from the

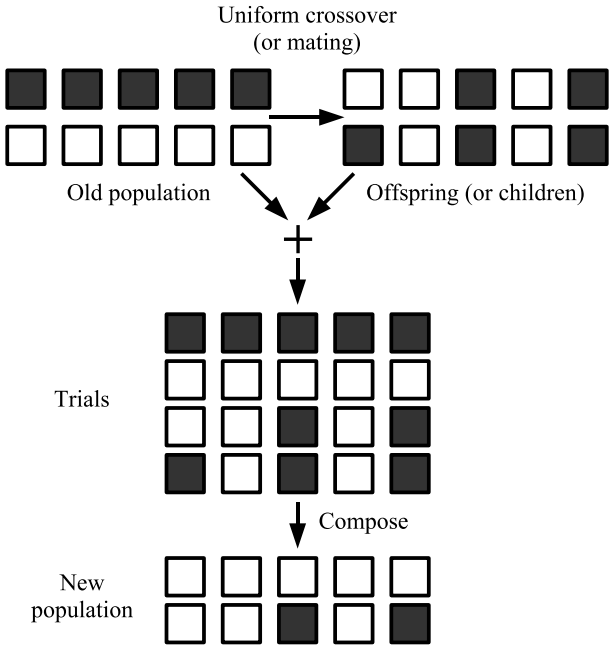


Figure 3.1: Principle of GA. The old population is consisted of parents. The parents means the parameter sets which are used in the composition of the next solutions. The next solutions, called offspring (or children), are formed from the parents using uniform crossover (or mating). Uniform crossover means that adjacent genes of a child are not probably selected from the same parent. In the next step, the old population and the offspring are joined together. In the compose step, the best of the trials form a new population. The cycle begins again, and the new population is used as the old population for a subsequent offspring.

second parent. The subsequent population is formed from the best individuals of the current populations and the offspring while the others are abandoned. This cycle is continued until certain criteria are met. The more detailed description of this procedure can be found from publication II.

The rules in the parent selection, the crossover mechanism, the formation of a new population, and in many other innerbuild parameters can be altered. Different modifications for GAs have been applied with success in x–ray analysis [56, 59–63]. However, the design of justified modifications to genetic algorithms is difficult since there is no general theory predicting the behaviour of the recurrent survival–of–the–fittest mechanism. One approach to design modifications is to study the problem in a specific case and modify the algorithm in such a way that it has better convergence properties for that specific problem. These modifications, however, can be problem dependent and therefore techniques capable of adapting to underlying problems are beneficial. One of the most fascinating class of modifications is linkage learning (or reducing) techniques [93–95] where the interparameter dependencies are found during the search procedure. This means that GA finds separable variables during the search and optimizes them independently of the others.

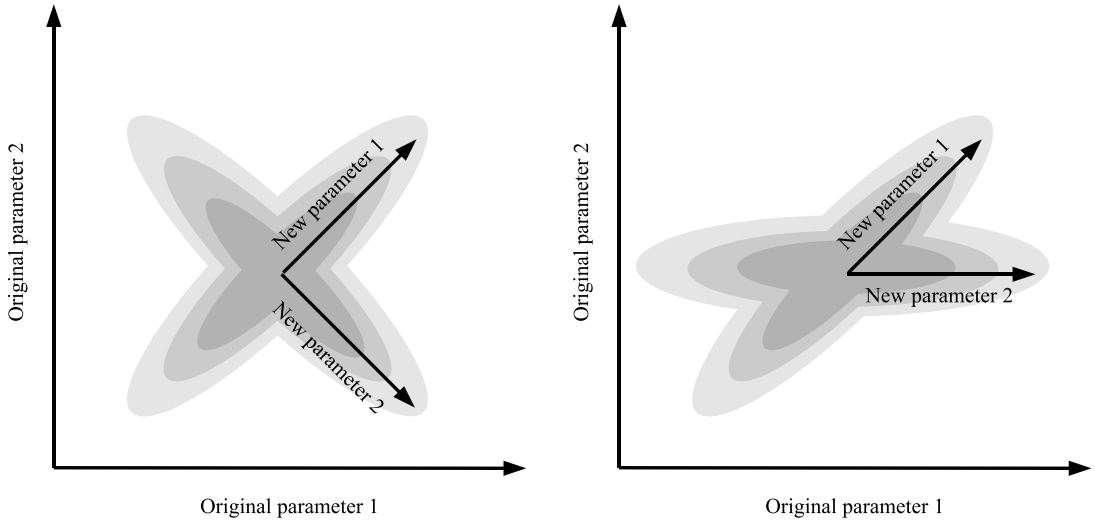


Figure 3.2: (a) Orthogonal problem domain presented by gray colours. The rotation of coordinates is performed to adapt the structure of the problem. (b) Nonorthogonal problem domain presented by gray colours. Nonorthogonal linear transformation is performed to adapt the structure of the problem.

The separation of the variables can be performed using the rotation of coordinates in an orthogonal search domain, as shown in figure 3.2(a). In general, a search space may not be orthogonal and it may be beneficial to use a nonorthogonal basis for the new coordinates, as illustrated in figure 3.2(b). The orthogonal and nonorthogonal linear transformations can also be advantageous in the mapping of similar original parameters to one parameter in the new search space.

The adaptation process for the structure learning can be exploited in GA using the principle shown in figure 3.3. The parents are transformed from the original search space to a new search space spanned by the new coordinate system using a separation matrix S. The child is composed using crossover in the new space and then transformed back to the original space using the inverse of the matrix. Subsequently, a new population is formed by removing the worst individuals, the separation matrix is updated, and the process starts again from the beginning.

The update step of S is here an essential part. One can determine the separation matrix from the basis of Hessian matrix or alternatively one can utilize statistical approach, for instance. In the stochastic approach, the sampling set can be based on a current population or all the individuals calculated during the optimization procedure. The sampling set can be used to calculate a suitable basis for the transformation with a method like principal component analysis which removes interparameter covariances of the sampling set by the rotation of coordinates. Another

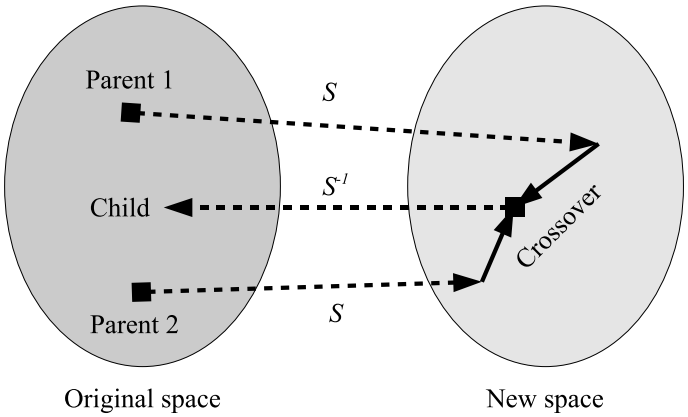


Figure 3.3: Schematics of the new concept in genetic algorithm. Parents 1 and 2 are transformed to the new search space spanned by the new coordinates using the separation matrix S. The crossover is performed in the new space and the composed child is transformed back to the original search space using the inverse of the separation matrix S^{-1}.

method could be independent component analysis which reduces mutual information between the parameters of the sampling set by a suitable nonorthogonal linear transformation [96].

3.1.2 Coordinate rotation techniques

Wyatt and Lipson introduced a coordinate rotation technique based on the Hessian matrix to reduce interparameter dependencies for real–valued problems in GA [94]. Similar discussion related to the coordinate rotation is presented also in reference [95] but the performance of the approach is not actively studied in XRR curve fitting.

The rotation of coordinates can be done in several ways but two important ones are given here. For the rotation, one needs to compute Hessian

$$H_{ij} = \left.\frac{\partial^2 F(\mathbf{p})}{\partial p_i \partial p_j}\right|_{\mathbf{p}}, \tag{3.1}$$

where F is the fitness function measuring the difference between a theoretical and the measured XRR curves and p_i and p_j are the components of the row vector $\mathbf{p}$ containing thickness, mass density and roughness data. The computational cost of the symmetrical Hessian matrix is approximately $2d^2$, where d is the dimensionality of the problem. Therefore techniques to reduce computation time due to expensive XRR curve calculations made for fitness value determination are required. One possibility is to reduce the number of points used in XRR curve computation thus reducing significantly the time required for Hessian matrix calculation. As an alter-

native technique for the computation time reduction, one can utilize the observations made during the search process in GA. Let us define a $N \times d$ sample matrix $\mathbf{P}$, where N represents the number of observations. One can compute the sample covariance matrix defined as

$$\mathbf{C} = (\mathbf{P} - \overline{\mathbf{P}})^T(\mathbf{P} - \overline{\mathbf{P}})/(N-1), \tag{3.2}$$

where $\overline{\mathbf{P}}$ is the mean matrix. For the rotation of coordinates, the computation of a new basis is required. The basis is the set of eigenvectors computed by using standard eigenvalue decomposition defined as

$$\mathbf{S} = \mathrm{Eig}(\mathbf{X}), \tag{3.3}$$

where Eig is the function determining the eigenvectors of the matrix $\mathbf{X}$. The argument $\mathbf{X}$ can be replaced by the Hessian or covariance matrices, and $\mathbf{S}$ is the matrix containing the eigenvectors of the new basis as the column vectors. The parameters in the new coordinates are obtained by

$$\mathbf{p}_{\mathrm{new}} = (\mathbf{S}^T \mathbf{p}_{\mathrm{original}}^T)^T. \tag{3.4}$$

The parameters are transformed back to the original coordinates by using the mixing matrix $\mathbf{M} = \mathbf{S}^{-1}$ in the mapping. The efficiency of GAs using the rotation of coordinates is studied in section 4.1.

3.1.3 Independent component analysis

The previous discussion concentrated on the use of Hessian and covariance matrices in the linear transformation. The methods based on these matrices produce an orthogonal basis but in general, a problem domain may be nonorthogonal. One efficient technique for the nonorthogonal transformation is independent component analysis (ICA) [96]. Whereas the covariance approach, called principal component analysis (PCA), transforms the new parameters to be uncorrelated, ICA transforms the original parameters to be independent, *i.e.*, the value in one new parameter does not give any information on the value in the other new parameter. On an intuitive level, this approach relies on the Central Limit Theorem which tells the sum of parameters from independent distributions tends toward a Gaussian distribution. By performing a suitable linear transformation, one can get independent components which are distributed as nongaussian way as possible.

Since the material researchers are typically unfamiliar with ICA, the efficiency of one ICA technique, called EFICA [97], is here compared to PCA in a demonstration purpose. The demonstration is based on a signal separation task where different ICA techniques are typically applied. Figure 3.4(a) shows three synthetic signals: A saw tooth signal, a sine wave and uniformly distributed noise. A 3×3 matrix

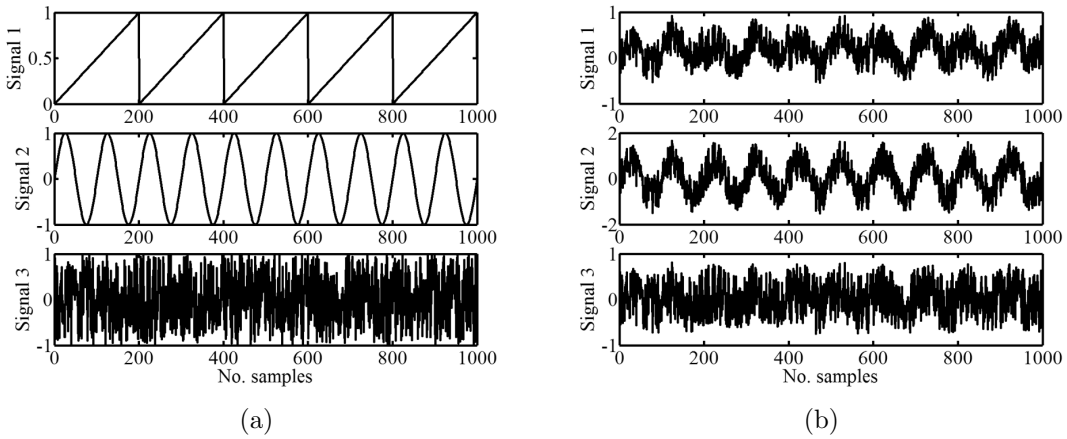


Figure 3.4: (a) Original signals. Signal 1 is the saw tooth signal, the second signal is calculated from the sine function and the third signal consists of random numbers from the uniform distribution. (b) Mixed signals.

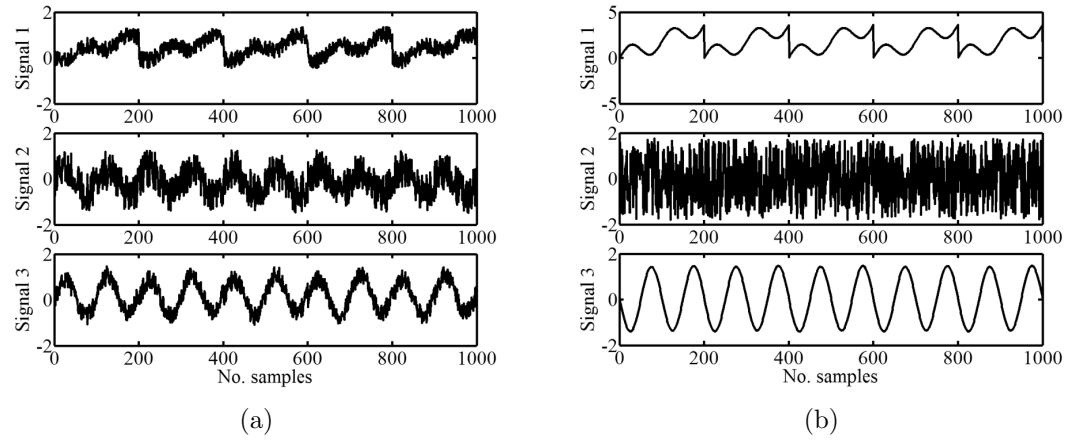


Figure 3.5: (a) Mixed signal separated by PCA (b) Mixed signal separated by EFICA.

consisting of random numbers was used to mix the original signals, as shown in figure 3.4(b). The mixed signals were separated using PCA and ICA, shown in figures 3.5(a) and 3.5(b), respectively.

One can observe that PCA has not succeeded to separate noise from the mixed signals. The separated signal 1 resembles the saw tooth signal although it has noise and an additional sine component. The separated signal 3 resembles the most the original sine wave but the signal is not clearly different from the separated signal 2. ICA, on the other hand, has separated noise and the sine wave while the saw tooth wave has false, but noiseless, reconstruction. If one does know the original signal, the results of ICA seem to be better separated than the signals separated by PCA.

Therefore the results suggest that ICA is a tempting method in general to be used to find linear dependencies between parameters of an input data. In other words, if the original parameters can be separated in the optimization, ICA is a better method for the separation task than PCA since the orthogonal transformation used in PCA is a special case of linear transformations. Therefore ICA can be used in the crossover phase of GA for the optimization of mutually independent data. ICA has, however, some disadvantages which can reduce its efficiency in the considered application:

- The variances (or scales) of the separated independent components cannot be determined. The false scales can cause problematic scaling of the original parameters in the inverse transformation.
- The separation matrix is solved in EFICA iteratively which means that the separation process is time consuming since each element of the $d \times d$ separation matrix needs processing. In other words, the computation time scales as $\mathcal{O}(d^2)$ which sets some practical limit for the dimensionality of the problem to be solved.
- The separation matrix is initialized before the iteration by random numbers which reduces the repeatability of the separation.
- The ICA algorithm may converge to a local optimum instead of a global optimum [98] which means unsuccessful separation.
- The separation of an input data may suffer from an insufficient number of observations. This is called as a finite sampling problem [99, 100].

These limitations together with the lack of efficient ICA algorithms and reasonable computational power may be the reason why GAs with ICA have been rarely studied although a few [101–103] have addressed that this approach is providing. In this thesis the efficiency of GA using ICA for XRR curve fitting is studied in section 4.2.

3.2 Novel x–ray reflectivity error analysis methods

3.2.1 Fitness function

A quantitative number representing the correspondence between the measurement and a theoretical curve is necessary in XRR analysis. A goodness–of–fit [104], here called as fitness (value), measures the difference between XRR curves, typically one curve being theoretical and the second being the measurement. In XRR measurements, photon counting causes Poisson noise and therefore a fitness measure denoted

here as

$$\chi^2 = \sum_{i=1}^{N} \frac{(\overline{x}_{i,\mathrm{fit}} - x_{i,\mathrm{meas}})^2}{\overline{x}_{i,\mathrm{fit}}}, \tag{3.5}$$

where $\overline{x}_{i,\mathrm{fit}}$ is the ith point of a theoretical XRR curve and $x_{i,\mathrm{meas}}$ is the measurement, is considered to be a good choice for a fitness value calculation. The fitting carried out by using χ^2 results, however, visually in a fit which is not matching with the measurements, as demonstrated in figure 3.6(a). Better agreement with the visual correspondence can be obtained using a fitness measure defined as

$$F = \frac{1}{N} \sum_{i=1}^{N} [\log x_{i,\mathrm{meas}} - \log \overline{x}_{i,\mathrm{fit}}]^2, \tag{3.6}$$

for instance. Figure 3.6(b) shows that F gives match between the fit and the measurement. The discussion presented here, however, shows that the selection of fitness measure in XRR curve fitting is not straightforward and requires careful attention.

3.2.2 Fitness landscape for fitness function analysis

The previous discussion shows that the selection of a fitness measure is problematic. The measure can be too sensitive meanwhile diminishing the convergence properties of a fitting algorithm, or vice versa. In other words, improperly selected fitness measure can cause a fitting procedure to approach 'needle in a haystack' problem which means that the exact solution has minimum of the fitness but in practice it cannot be found.

The properties of fitness functions can be studied by using a tool called fitness landscape [105, 106]. A fitness landscape resembles an ordinary landscape in nature where the altitude of the land represents the fitness value as a function of location. It is worth recalling here that the search algorithm performs the optimization blindly which means that a 'mountain' decreases the possibility to find the next 'valley' behind that 'mountain'. To get a better picture of the challenge, one can consider a case where a blind man is searching a valley in certain altitude from a mountain chain but cannot find other valleys or differentiate one valley from another. Therefore fitness landscapes are valuable tools since they provide direct visual observation of a search domain for a researcher. The researcher can study the properties of a fitness landscape with different fitness measures and make qualitative conclusions about the feasibility of measures for the given problem.

Fitness landscapes are practical if the search domain contains no more than two parameters. Even the simplest XRR curve fitting problems have more parameters to be fitted and the use of fitness landscapes is difficult without the reduction of the dimensionality of the problem. The dimensionality reduction can be performed

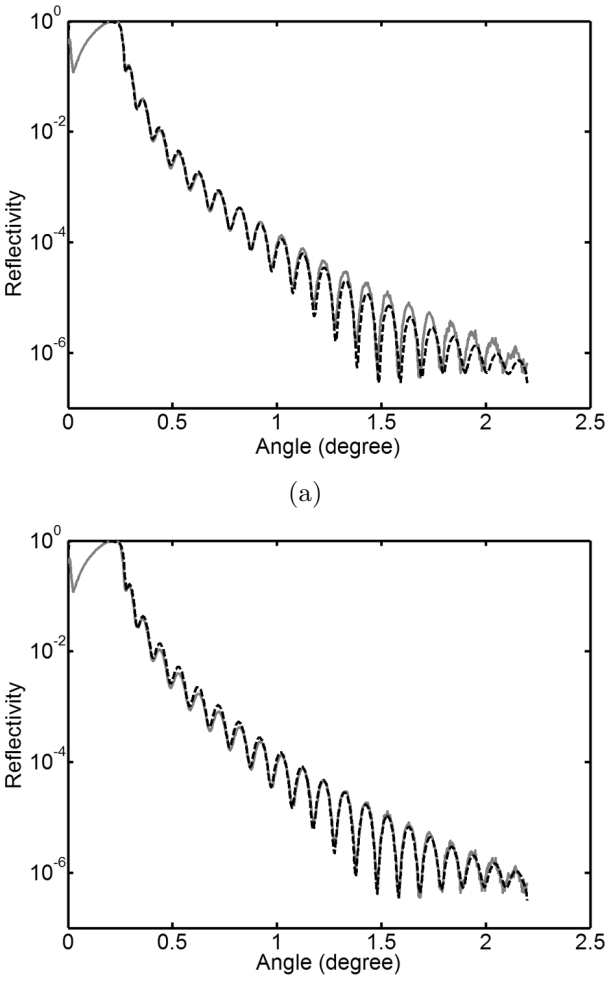


Figure 3.6: Measured (gray solid line) and fitted (black dashed line) XRR curves of an ALD AlO sample. The black dashed curve is fitted using the (a) χ^2 and (b) F measures [publ. V].

by modelling the XRR properties of samples deposited by ALD using the process parameters, *i.e.*, the number of cycles and the growth temperature, as the search parameters. Possible grooves in fitness landscapes tells that XRR is not sensitive to certain combinations of ALD process parameters, *i.e.*, if fitnesses are nearly equal, then samples fabricated with certain combinations of ALD parameters are indistinguishable for XRR. This issue based on the presented scope is studied in section 4.3.

3.2.3 Fitness and Poisson noise

One way to understand inaccurate parameter determination in XRR is to consider the consequence of Poisson noise on fitness values. Consider the case when one adds Poisson noise to a theoretical curve and calculates the fitness between the original and the noisy curve. In this case, the calculated fitness value varies since the exact form of noise is unknown. However, by repeating the fitness calculation between the original and new noisy curves, one can get a fitness distribution.

Similarly, Poisson noise alters the fitness between measurements even when the measurement is carried out on the same sample. In practice this means that the minimum fitness value for an arbitrary measurement is unknown but the value belongs to a certain distribution. Since the exact minimum fitness value is unknown, the determination whether the found solution is correct or not is based on heuristics. In other words, the fitting procedure is restarted if a researcher is not satisfied with the visual correspondence between a theoretical XRR curve and the measurement.

The introduced problem can be circumvented by using simulated XRR curves and their fitness values to study the probability of the existence of the minimum fitness at the given location. In practice one approaches the problem by studying the question whether the assumed exact solution is the best fit for the noisy curve or not. Since Poisson noise is a stochastic process, the question should be studied using an appropriate test method. One can use statistical pairwise comparison test between the assumed exact solution and an alternative curve. If the probability of the exact solution having better fitness than the alternative curve decreases below a certain limit, one cannot surely say that the exact solution has the minimum fitness.

The appropriate test method for pairwise comparison is the p–value test which tests a given hypothesis. The p–value is a ratio of the number of events agreeing with the hypothesis divided by the number of all events. If the p–value is greater than a significance level, called α, the test shows that there is not enough evidence against the hypothesis but says nothing about the validity of the hypothesis itself. With this condition, the test gives a result from an acceptance region. If the p–value is less than α, the test gives a result from a rejection region which means that the hypothesis is rejected. In practice, one should define some α and find a parameter set where the p–value equals α. That parameter set gives an uncertainty region caused by Poisson noise where one cannot say that the found solution is the exact solution. The presented scope using the p–value test is utilized in section 4.4.

3.2.4 Poisson noise and nonideal fit

The previous discussion assumed that the found solution is the exact solution which is not strictly speaking true in general. Therefore an approach taking into account

the effect of a nonideal fit and Poisson noise on fitness value is necessary. Using the definition presented in equation 3.6 one can approach the problem by identifying two important scenarios in the fitting process:

1. The measurement x_{meas} contains noise but a noiseless theoretical XRR curve $\overline{x}_{\text{opt}}$ is the exact solution to the measurement. In such a case the fitness represents the fitness of noise which is defined as:

$$\begin{aligned} F_{\text{noise}} = & \quad F(x_{\text{meas}}, \overline{x}_{\text{opt}}) \\ = & \quad N^{-1} \| \log x_{\text{meas}} - \log \overline{x}_{\text{opt}} \|_2^2. \end{aligned} \tag{3.7}$$

   Note that F_{noise} represents the minimum fitness which can be obtained. The value of F_{noise} can be simulated by approximating that $x_{\text{meas}} \approx \overline{x}_{\text{opt}} +$ Poisson noise $= x_{\text{opt}}$ which yields

$$\begin{aligned} F(x_{\text{meas}}, \overline{x}_{\text{opt}}) = & \quad N^{-1} \| \log x_{\text{meas}} - \log \overline{x}_{\text{opt}} \|_2^2 \\ \approx & \quad N^{-1} \| \log x_{\text{opt}} - \log \overline{x}_{\text{opt}} \|_2^2 \\ = & \quad F(x_{\text{opt}}, \overline{x}_{\text{opt}}). \end{aligned} \tag{3.8}$$

2. The measured curve contains no noise or noise is negligible. This is the case when the signal–to–noise ratio is high, especially when high–power synchrotron radiation is used as an x–ray source. In this case $x_{\text{meas}} \approx \overline{x}_{\text{meas}} = \overline{x}_{\text{opt}}$ and now the fitness can written as

$$\begin{aligned} F(x_{\text{meas}}, \overline{x}_{\text{fit}}) \approx & \quad N^{-1} \| \log \overline{x}_{\text{meas}} - \log \overline{x}_{\text{fit}} \|_2^2 \\ = & \quad N^{-1} \| \log \overline{x}_{\text{opt}} - \log \overline{x}_{\text{fit}} \|_2^2 \\ = & \quad F(\overline{x}_{\text{opt}}, \overline{x}_{\text{fit}}) \end{aligned} \tag{3.9}$$

   where $\overline{x}_{\text{meas}}$ represents the measured curve without noise, $\overline{x}_{\text{fit}}$ is the fitted theoretical curve and now the fitness is totally dependent on the nonideal fit.

These two special cases play a keyrole in the determination of error including noise and nonideal fit. In section 4.5, it is shown that the original fitness is approximately the sum of the right sides of equations 3.8 and 3.9 when the upper limit for an error is determined.

4 Results and discussion

This chapter discusses GAs using linear transformation techniques in XRR curve fitting and the novel error analysis methods developed for XRR analysis. The work presented here is based on the text introduced in the previous chapter. Section 4.1 and 4.2 present the novel GAs developed in publications I and II, respectively. The present the novel error analysis methods studied in publications III, IV and V are presented in sections 4.3–and 4.5.

4.1 Genetic algorithms using rotation of coordinates

Publication I introduced two novel GAs based on the coordinate rotation techniques for XRR analysis. The first algorithm, called HGA, utilized a numerically computed Hessian matrix while the second algorithm, called CovGA, utilized interparameter covariance matrix in the calculation of a new basis.

The basis of the coordinate rotation in HGA was the set of the eigenvectors of the Hessian matrix in the vicinity of the best individual and the basis in CovGA was the set of the eigenvectors of interparameter covariance matrix calculated from population values. The new generation was composed in rotated coordinates and after the crossover, the offspring was transformed back to the original coordinates where the fitness values were calculated. Additionally, the novel scheme reshaping the probability distribution of the parameters, called nonlinear–fitness–space–structure adaptation (NL–FSSA), was introduced to enhance the convergence properties of CovGA and HGA. Every cycle in the GAs included the following steps:

1. The initialization of a population.
2. Individuals are moved in XGA [56].
3. A certain number of randomly selected individuals are selected to mate with the best individual.
4. NL–FSSA is performed.
5. The rotation of coordinates is performed. CGA uses identity transformation.

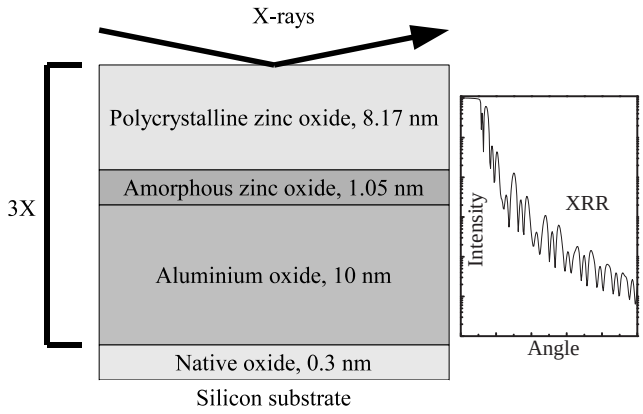


Figure 4.1: Schematic structure of the modelled nanolaminate and the theoretical x-ray reflectivity curve used in the performance tests [publ. I].

6. A local search is performed for the offspring in rotated coordinates.
7. The new individuals are transformed back to the original coordinates.
8. The new population is constructed.

The performance of the algorithms was tested by fitting randomly selected trials to the XRR curve based on a realistic model of a three–period ZnO/AlO nanolaminate shown in figure 4.1.

CovGA and HGA were compared to the classical GA and to eXtended GA (XGA) [56]. Figure 4.2(a) shows the convergence curves of 50 populations and that novel algorithms had fewer stagnated populations than conventional methods as a function of cycles. Note also that CovGA minimizes averagely the fitness better than HGA. Zoomed median curves in figure 4.2(b) confirm qualitatively the difference between the previous observations.

Figure 4.3 shows the fraction of solutions (%) below a certain fitness after 50 cycles. Results show that the robustness of convergence is slightly better for XGA than for CGA in the studied problem, and that the novel algorithms exhibit notably better convergence statistics than the other methods. The effect of switching NL–FSSA to the performance of novel algorithms was studied. The performance of the novel algorithms without NL–FSSA diminished but however, the novel methods with coordinate rotation were still clearly better than CGA or XGA.

To conclude, CovGA, had the best fitting properties in the studied case even when NL–FSSA was excluded. The performance of CovGA was based on the rotation of coordinates using the eigenvectors of the covariance matrix as the basis. It was shown later in publication II that this rotation of coordinates can be further improved by generalizing the coordinate rotation technique to a suitable linear trans-

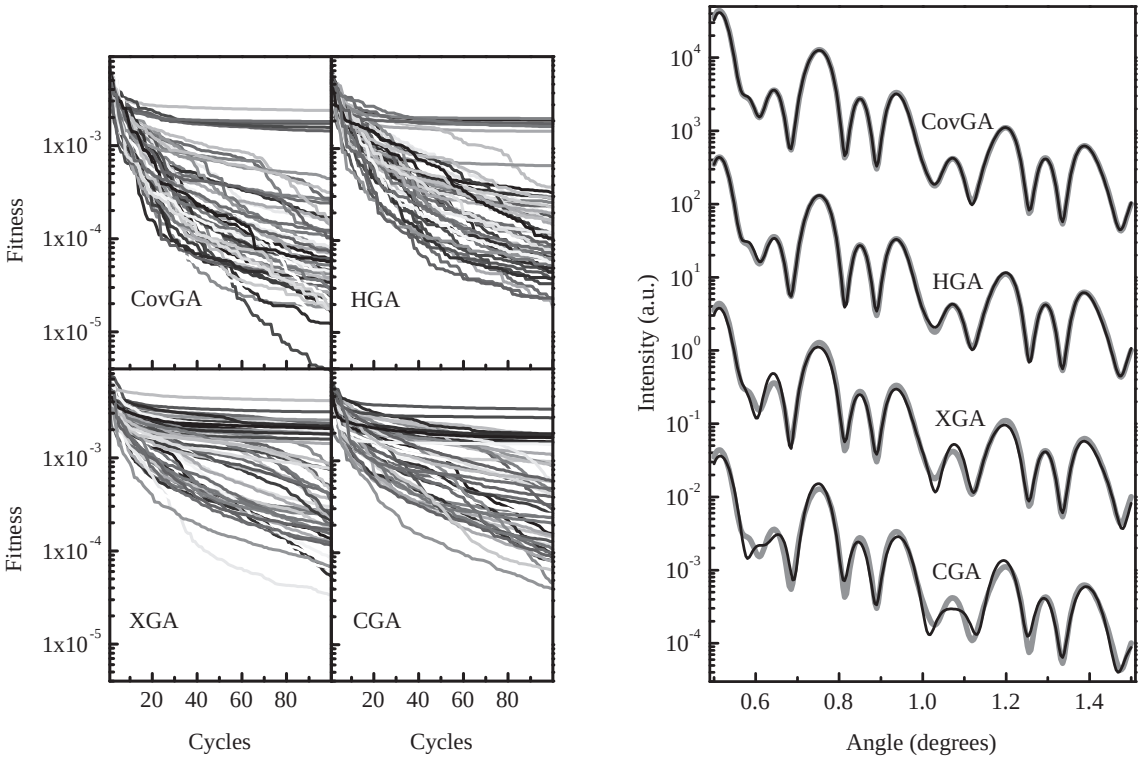


Figure 4.2: (a) Fitness of the best individual in 50 different populations as a function of cycles. Each population has its own line in greyscale. (b) Fitted median curves. The simulated XRR curves based on the model are plotted with dark grey lines and the median fits with solid black lines. The curves have a vertical offset for clarity. [publ. I].

formation in XRR curve fitting. The further development of NL–FSSA will be done in future.

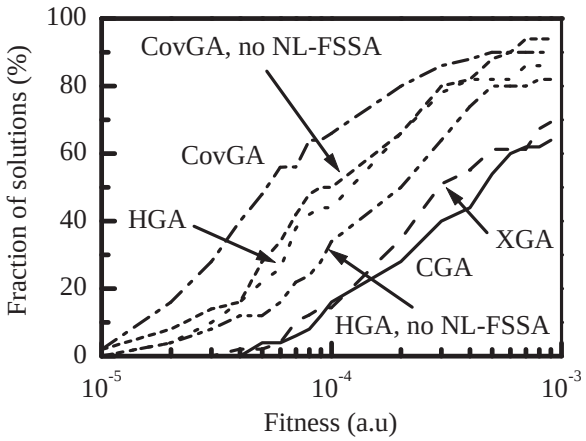


Figure 4.3: Fraction of solutions (%) below a certain fitness. The results of CovGA and HGA without NL–FSSA are also shown [publ. I].

4.2 Genetic algorithm using independent component analysis

The success of the covariance based statistical approach as a fitting method in publication I suggested that stochastic approaches may be good choice for the parameter separation. In publication II, a more sophisticated statistical technique than the covariance approach, called independent component analysis (ICA), was used for the parameter separation and the performance of the novel GA was studied.

An efficient FastICA (EFICA) method [97] was applied for the computation of the separation matrix. Its advantages are robustness with generalized Gaussian distribution and it is fast. In the calculation, all the points computed during the search procedure were used as the sampling set in PCA and ICA to reduce possible finite sampling problems. The implementation of the novel GA using ICA (ICAGA) utilized the same features as CGA and on PCAGA, where PCAGA was the simplified CovGA. The only separating factor between all the studied algorithms was the linear transformation technique. Identity transformation was used in CGA. Figure 4.4 shows the implementation used in the GAs. Cycles in the GAs included the following steps:

1. The best individual in the population is selected to mate with 12 randomly selected individuals.
2. PCA or ICA is performed to compute mixing and demixing matrices. Identity matrix is used in CGA.
3. A crossover operator utilizing linear transformation is used in the creation of new individuals.
4. The back transformation is used for the offspring.
5. A fast local random search is used instead of mutation for the children
6. The new trial population consisting 30 members is composed from the best individuals of the current population and the offspring.

The performance of the algorithms was tested using a multiperiodic ZnO/AlO layer model shown in figure 4.5. A target curve was calculated using the layer structure and random trial curves were fitted to the target curve.

The result was that CGA had worse performance than PCAGA and ICAGA when the number of periods is more than one, as shown in figure 4.6(a). ICAGA was outperforming others when the number of periods was greater than three. Figure 4.6(b) shows the trial curves of the median fits in the 15–period case for the algorithms. The curve fitted by ICAGA is following the target curve clearly better than the

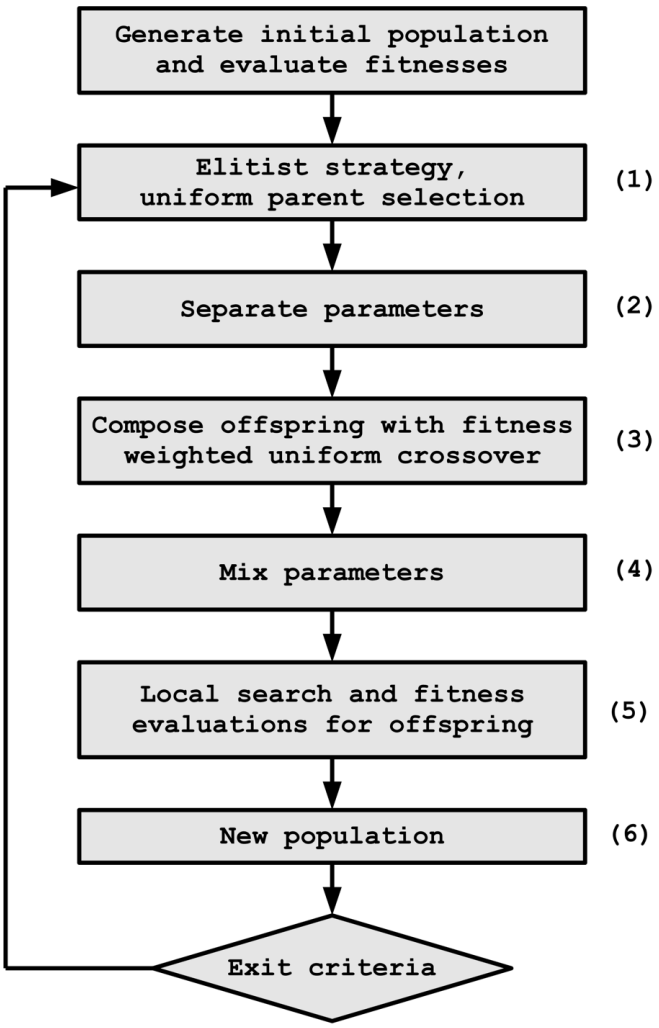


Figure 4.4: Flowchart of GAs [publ. II].

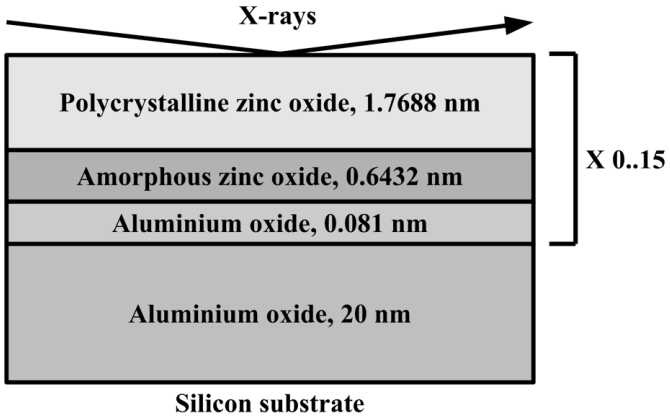
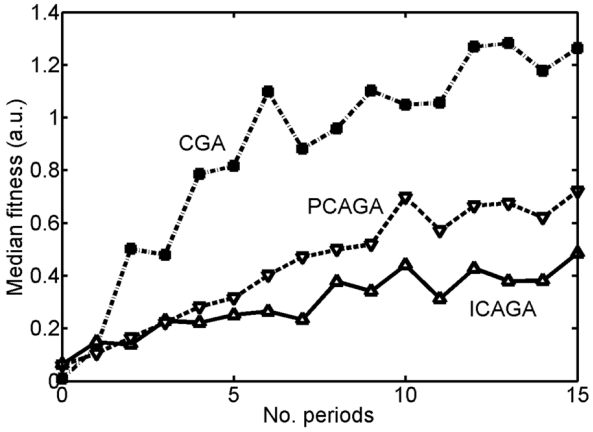
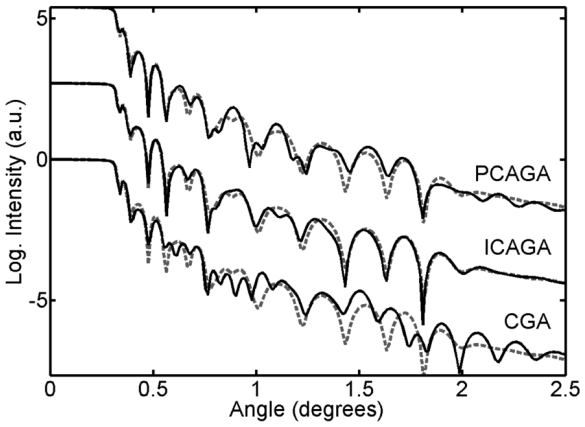


Figure 4.5: Schematic structure of the modelled nanolaminate with varying number of thin ZnO/AlO-periods [publ. II].



(a)



(b)

Figure 4.6: (a) Median fitnesses of CGA, PCAGA and ICAGA as a function of period number. The median fitnesses are based on 25 fits in each datapoint. (b) The median XRR fits in the 15–period case (black solid line) are based on 25 simulations. The target curves are in gray dotted lines. The curves have a vertical offset for clarity [publ. II].

curves fitted by other methods which confirms the differences between the fitness values. Thus it can be concluded that ICAGA has the best capability to extend the analysis based on Parratt's formalism to multiperiodic layer structures.

The disadvantage of the proposed ICAGA is, however, the increased time consumption shown in figure 4.7. This was a consequence of quadratically growing size of the separation matrix which is iteratively solved by a fixed-point algorithm in EFICA. This sets a practical limit for the problem dimensionality due to the increasing computational requirement but the increasing time consumption is independent of curve computations. Therefore the relative difference in time consumption between the

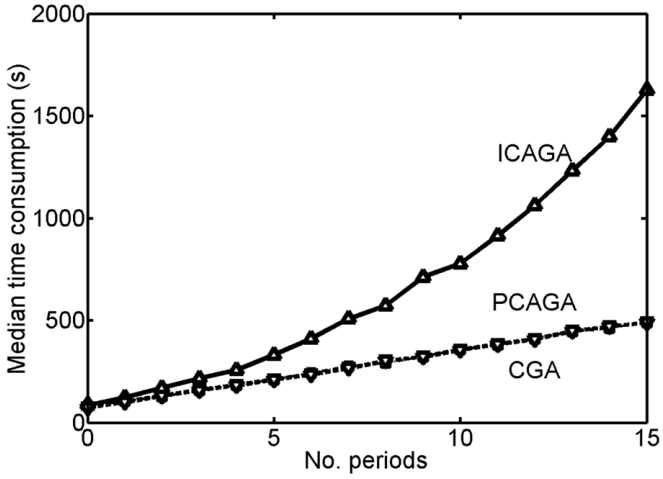


Figure 4.7: Median time consumption of CGA, PCAGA and ICAGA when a different number of periods is applied. The median times are based on 25 fits done for each datapoint [publ. II].

new and other algorithms is decreased when the population size or the number of datapoints used in one XRR curve calculation is increased. However, the present results suggest that techniques for the dimensionality reduction of the search space are worth to be studied in the future.

4.3 Crosserror between the mass density and surface roughness

Nonuniqueness or ambiguity of the determined parameters is a fundamental interest in XRR curve fitting problems but the topic is rarely studied [58]. This problem was approached in publication III using analysis based on fitness landscape simulations where the height in a simulated map represents the fitness value as a function of the parameters. The studied material was ALD AlO and the layer model is shown in figure 4.8.

Since the use of fitness landscapes is difficult if there are more than two dimensions, the dimensionality reduction of the parameter space was performed by modelling the empirical XRR properties of the layer structure by feedforward neural neural networks. The use of these neural networks provided the modelling without *a priori* knowledge of the order of nonlinearities. The modelling was performed using XRR data measured from ALD AlO samples deposited by Beneq TFS 500 reactor. Figure 4.9 shows the principles of the approach used in this study for the dimensionality reduction. The original six XRR parameters were modelled by the two ALD process parameters, the number of cycles and the deposition temperature. Other parameters were fixed during the modelling since no systematic differences in the parameters were found between the fabricated samples. The observation indicates that the upper most layer dominates the features of XRR curves while the properties of the underneath layers differ only slightly between the samples.

The fitness landscapes were employed to study sensitivity properties of the fitness function used in XRR curve fitting. The applied approach allows to clear out such combinations of ALD process parameters which produce indistinguishable samples for XRR measurements. The fitness landscapes were computed choosing first an optimal solution as an origin for a certain fitness landscape. The number of cycles and the deposition temperature were varied around the origin. The fitness value for the varied parameter was calculated using the following steps:

1. Select an optimal solution as the origin of the fitness landscape.

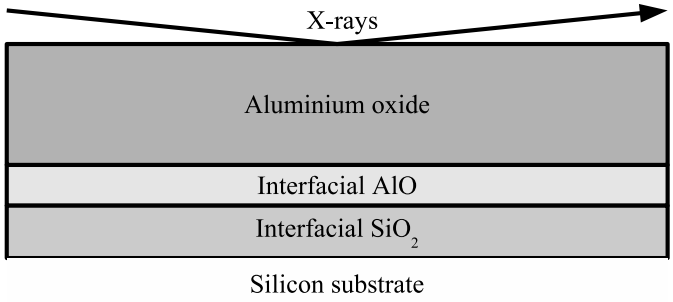


Figure 4.8: Schematic layer model used in XRR curve calculation [publ. III].

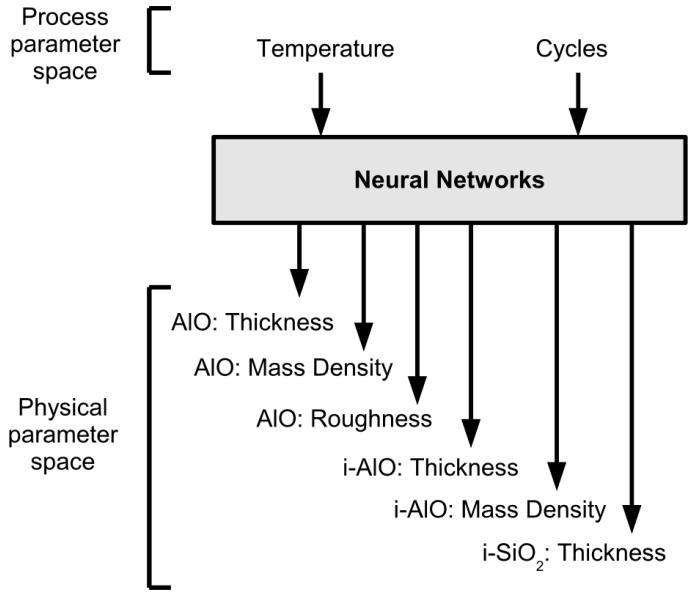


Figure 4.9: Dimensionality reduction of the problem by neural networks. The two ALD process parameters are input parameters for the neural networks. The six output variables are in the physical parameter space. Abbreviations i–AlO and i–SiO_2 are interfacial AlO and SiO_2, respectively.

2. Map the optimal solution to the physical parameters.
3. Simulate a XRR curve based on the physical parameters.
4. Select a point near the optimal solution.
5. Map the selected point to the physical parameters.
6. Simulate a new XRR curve based on the new physical parameters.
7. Calculate fitness between the new and the original curve.
8. Go to the step 4 until all necessary points are computed.
9. Draw the fitness landscape.

All the points calculated using the above described approach were collected to a fitness landscape. The procedure was repeated several times which produced multiple fitness landscapes. Further analysis was based on the calculated fitness landscapes, one of them shown in figure 4.10.

It was observed that the calculated fitness landscapes contain grooves where minima of fitness values are located. The shape of the groove was different for each fitness landscape. The grooves were projected using the neural networks to the physical parameters. Figure 4.11 shows the projected grooves as a function of the origins of fitness landscapes.

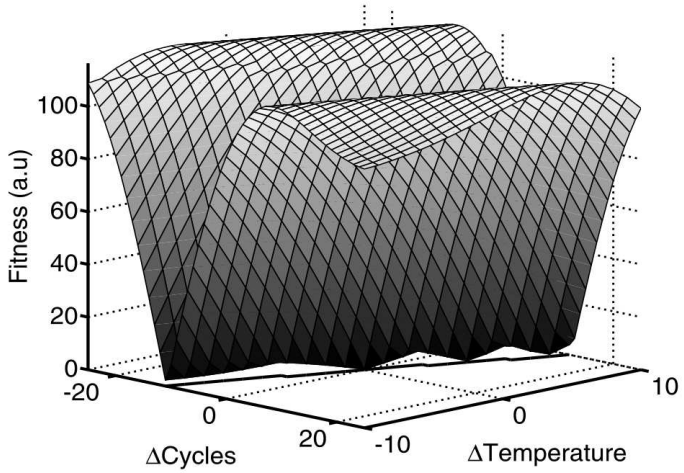


Figure 4.10: Fitness landscape as a function of the ALD process parameters. The origin is at temperature 300°C and 500 cycles on the process parameter space. The black line in the (ΔCycles, ΔTemperature)-plane is the projection of the groove in the fitness landscape [publ. III].

By taking the slope of the line with 300 cycles and 150 °C, one obtains

$$\Delta\text{Roughness} \approx 0.11 \times \Delta\text{Mass Density}, \tag{4.1}$$

where roughness is expressed in nanometres and mass density is in g cm^{-3}. Although this result is strictly limited to the studied case, it shows objectively that the mass

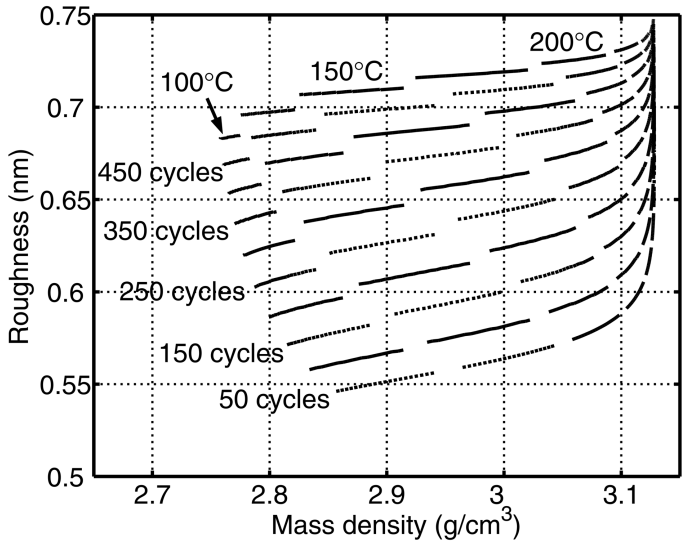


Figure 4.11: Projected grooves in the (mass density, roughness)-plane. The process parameters show the origins of fitness landscapes [publ. III].

density and the surface roughness are not independently determined factors in XRR analysis, as it will be shown again with other methods in publications IV and V.

4.4 Poisson noise limited accuracy in x–ray reflectivity analysis

The Poisson noise limited accuracy in XRR analysis was studied in publication IV using a model of a single aluminium oxide layer on silicon shown in figure 4.12.

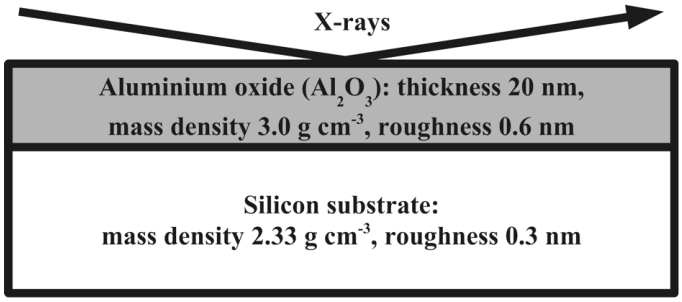


Figure 4.12: Schematic structure of an Al_2O_3 layer on silicon [publ. IV].

A hypothesis which argued that

H_0: *Other than the exact solution gives the best fitness*

was examined with the p–value test using a significance level of $\alpha = 0.01$. Simulations were performed for a fit instead of a measurement since the exact error caused by noise cannot be determined from the measurement. The fitness calculation for the exact solution was performed with the following procedure:

1. Compute a theoretical XRR curve $\overline{x}_{\text{fit}}$.
2. Set a new seed number for a random number generator.
3. Create a noisy XRR curve x_{fit} from the theoretical XRR curve $\overline{x}_{\text{fit}}$ by adding synthetic Poisson noise to it.
4. Calculate a fitness $F(x_{\text{fit}}, \overline{x}_{\text{fit}})$ using equation 3.6 between the theoretical and noisy XRR curves.
5. Go to the step 2 until a sufficient number of fitness values is calculated.

After this simulation step, the hypothesis can be performed using the following procedure:

1. Suppose the found solution $\overline{x}_{\text{fit}}$ is the exact solution.
2. Create noisy curves x_{fit} containing Poisson noise at the exact solution.
3. Calculate fitness distribution consisting of $F(x_{\text{fit}}, \overline{x}_{\text{fit}})$'s between the noisy curves x_{fit}'s and the noiseless curve $\overline{x}_{\text{fit}}$ using the above mentioned approach.
4. Move away from the exact solution, calculate a new theoretical XRR curve $\overline{x}_{\text{trial}}$ at that point.
5. Calculate new fitnesses $F(x_{\text{fit}}, \overline{x}_{\text{trial}})$'s between the new XRR curve $\overline{x}_{\text{trial}}$ and the old simulated noisy XRR curves x_{fit}'s.
6. Compare pairwise the new fitnesses $F(x_{\text{fit}}, \overline{x}_{\text{trial}})$'s and the old fitnesses $F(x_{\text{fit}}, \overline{x}_{\text{fit}})$'s.
7. Go to the step 4 if the new fitnesses are better than the old fitnesses with a certain probability.
8. One point of the confidence limit is found.

Note that this description shows a general approach for a confidence limit determination for the accuracy. It does not, however, consider practical implementation for the determination procedure. In publication IV, the confidence limits were determined at (mass density, surface roughness), (thickness, surface roughness) and (thickness, mass density) planes. A trial solution which was assumed initially to be the exact solution, was varied radially at the selected plane until the confidence limit based on the hypothesis testing was met. This implementation was selected to reduce computation time. Monotonic behaviour of the p–value was assumed as a function of the distance from the assumed exact solution. The determination of the confidence limit, where $p = \alpha$, for XRR parameters was done as follows:

1. Define the exact parameter set as $\mathbf{p} = [t, \rho, \sigma]$, where t is the layer thickness, ρ is the mass density and σ is the surface roughness of the layer.
2. Calculate a noiseless XRR curve at $\mathbf{p}$.
3. Add Poisson noise to the noiseless curve $N = 1000$ times and save each curve.
4. Compute fitness between the N noisy curves and the noiseless curve. Save fitness values in a vector component F_i^{exact}, where $i = 1, 2, 3 \ldots N$.
5. Select the plane, for instance the (mass density, surface roughness) plane, where to determine confidence limits.
6. Define a search direction $\mathbf{d}(\theta)$, where $\theta \in [0°, 360°]$, for instance

$$\mathbf{d}(\theta) = \mathbf{p} \begin{bmatrix} 0 & 0 & 0 \\ 0 & \cos(\theta) & 0 \\ 0 & 0 & \sin(\theta) \end{bmatrix} \tag{4.2}$$

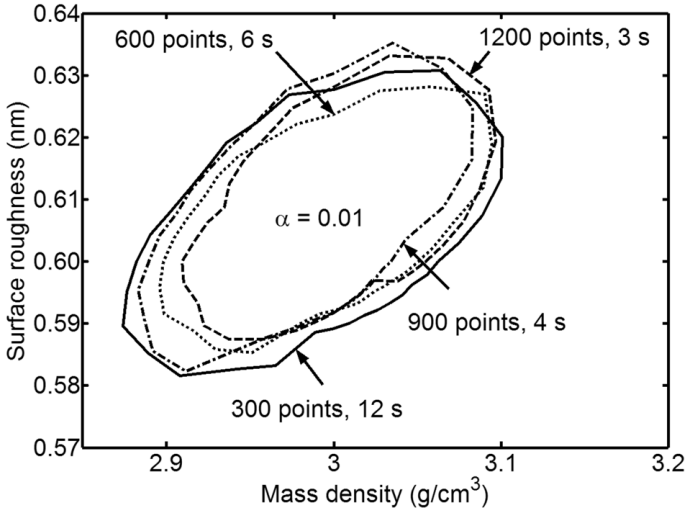


Figure 4.13: Confidence limits with $\alpha = 0.01$ when the measurement time is constant with altering averaging time [publ. IV].

(a) For each θ, find r such that $p(r) \approx \alpha$:

i. Calculate a XRR curve with the parameters $\mathbf{p} + r\mathbf{d}(\theta)$.

ii. Calculate its fitness with the previously mentioned exact XRR curve containing noise N times. Save the fitness values to a vector component F_i, where $i = 1, 2, 3 \ldots N$.

iii. Calculate

$$p = \frac{1}{N} \sum_{i=1}^{N} (F_i < F_i^{\text{exact}}). \tag{4.3}$$

The condition $(F_i < F_i^{\text{exact}})$ gives one if satisfied, otherwise zero.

iv. If $p \approx \alpha$, save $\mathbf{p}_{\text{contour}}(\theta) = \mathbf{p} + r\mathbf{d}(\theta)$.

7. Draw a closed contour $\mathbf{p}_{\text{contour}}$ in the selected plane.

Using the presented implementation, the confidence limits separating rejection (outside) and acceptance (inside) regions of the hypothesis, were determined. The determined confidence limit gave the fundamental limit for the accuracy in XRR analysis. The limited accuracy was a consequence of Poisson noise at a certain time which means that the confidence limit represents the lower limit of the error made in x-ray reflectivity analysis.

Here, two particular cases are discussed. In figure 4.13 the confidence limits were determined from parameters obtained from a measurement. One can observe that the relative lower error in the mass density is

$$\frac{3.1 - 2.85}{3.0} \times 100\% \approx 8.3\% \tag{4.4}$$

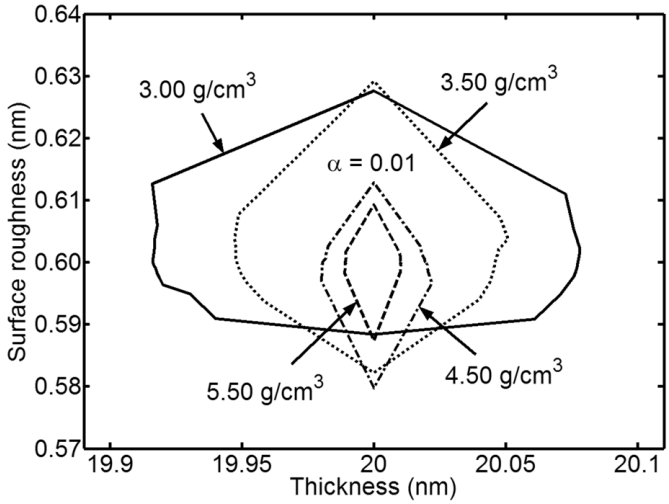


Figure 4.14: Confidence limits with $\alpha = 0.01$ as a function of mass density [publ. IV].

and is approximately equal to

$$\frac{0.63 - 0.58}{0.6} \times 100\% \approx 8.3\% \tag{4.5}$$

which is the relative lower error of the surface roughness.

One can conclude that errors of the mass density and the surface roughness are at the same level thus indicating possible crossinteraction. It is worth noticing here that the direction of principal axes of the confidence region are nonparaxial which is the evidence for the crossinteraction observed also in publication III. Similar nonparaxial shapes of the determined regions will be shown also in the next section presenting the results of publication V.

Note that the relative lower error of thickness in figure 4.14 is

$$\frac{20.09 - 19.91}{20} \times 100\% \approx 0.9\% \tag{4.6}$$

for ALD AlO with the mass density of 3.00 g cm^{-3} but the accuracy is better for materials having a higher mass density.

4.5 Error limit determination in x–ray reflectivity analysis

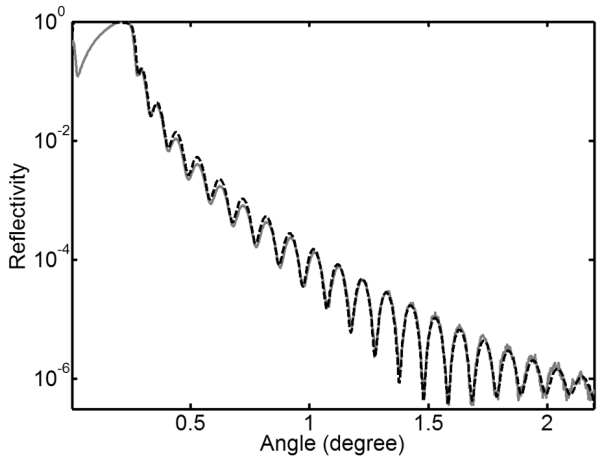
It was shown in publication IV that Poisson noise is an important factor which can cause a numerical optimum being elsewhere than the exact solution. However, the

study assumed that the fit equals the exact solution which is strictly speaking not true. A more advanced method taking into account also the separate location of the exact and the fitted solution was developed in publication V. The error limits were determined using a measurement and the fit shown in figure 4.15(a). The determined layer structure is shown in figure 4.15(b).

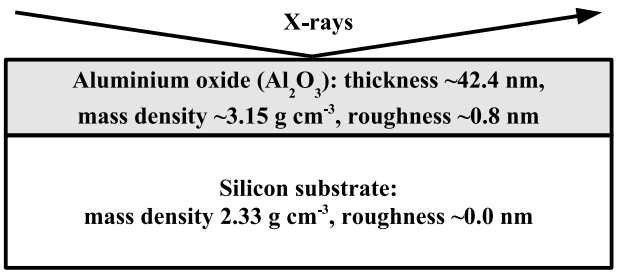
It was discussed in section 3.2.4 that the original fitness is the sum of the fitness of noise and the fitness of nonideal fit. Here the separation is shown in brief form. Let us denote the original fitness as

$$\begin{aligned} F(x_{\mathrm{meas}}, \overline{x}_{\mathrm{fit}}) = \; & N^{-1} \| \log x_{\mathrm{meas}} - \log \overline{x}_{\mathrm{fit}} \|_2^2 \\ = \; & N^{-1} \| (\log x_{\mathrm{meas}} - \log \overline{x}_{\mathrm{opt}}) \\ & + (\log \overline{x}_{\mathrm{opt}} - \log \overline{x}_{\mathrm{fit}}) \|_2^2 \\ = \; & N^{-1} \| e_{\mathrm{min}} + e_{\mathrm{res}} \|_2^2, \end{aligned} \tag{4.7}$$

where $e_{\mathrm{min}} = \log x_{\mathrm{meas}} - \log \overline{x}_{\mathrm{opt}}$ and the residual term $e_{\mathrm{res}} = \log \overline{x}_{\mathrm{opt}} - \log \overline{x}_{\mathrm{fit}}$ between the noiseless curves. Note here that e_{min} and e_{res} are the error components



(a)



(b)

Figure 4.15: (a) Measured (gray solid line) and fitted (black dashed line) XRR curves of an ALD Al_2O_3 sample. (b) Fitted layer model [publ. V].

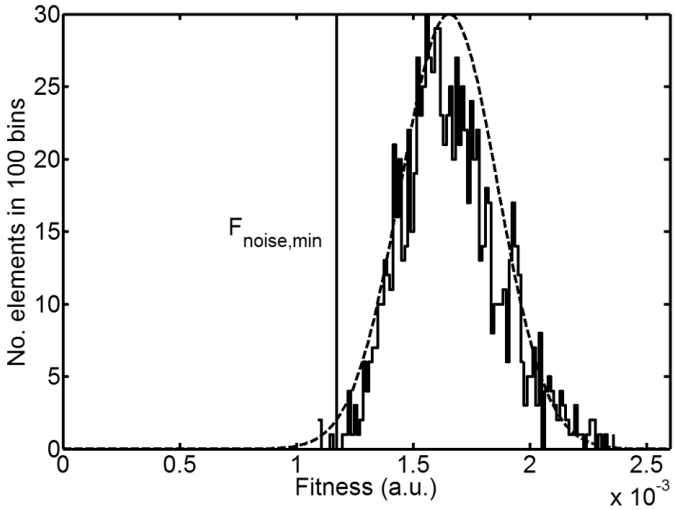


Figure 4.16: Fitness distribution of Poisson noise and the minimum of the fitness of noise [publ. V].

of the fitness of noise and the nonideal fit, respectively. Now the original fitness can be separated into three parts:

$$\begin{aligned} F(x_{\text{meas}}, \overline{x}_{\text{fit}}) = & \quad N^{-1}\|e_{\min} + e_{\text{res}}\|_2^2 \\ = & \quad N^{-1}\|e_{\min}\|_2^2 + N^{-1}\|e_{\text{res}}\|_2^2 + 2N^{-1} e_{\min} \cdot e_{\text{res}} \\ = & \quad F(x_{\text{opt}}, \overline{x}_{\text{opt}}) + F(\overline{x}_{\text{opt}}, \overline{x}_{\text{fit}}) + 2N^{-1} e_{\min} \cdot e_{\text{res}} \end{aligned} \tag{4.8}$$

Since $\overline{x}_{\text{opt}}$ and x_{opt} are unknown, the original fitness reduces to

$$\begin{aligned} F(x_{\text{meas}}, \overline{x}_{\text{fit}}) = & \quad F(x, \overline{x}) + F(\overline{x}, \overline{x}_{\text{fit}}) + 2N^{-1} e_{\min} \cdot e_{\text{res}} \\ \geq & \quad F(x, \overline{x}) + F(\overline{x}, \overline{x}_{\text{fit}}) - 2|N^{-1} e_{\min} \cdot e_{\text{res}}|. \end{aligned} \tag{4.9}$$

where $\overline{x}$ is a curve to be searched and x can be simulated. The inequality can be written in the form

$$F(\overline{x}, \overline{x}_{\text{fit}}) \leq F(x_{\text{meas}}, \overline{x}_{\text{fit}}) - F(x, \overline{x}) + 2|N^{-1} e_{\min} \cdot e_{\text{res}}|. \tag{4.10}$$

It was shown in publication V that $|N^{-1} e_{\min} \cdot e_{\text{res}}| \to 0$ for sufficiently large N and the number of detected photons. On the other hand, $F(x, \overline{x})$ was approximated by a separate simulation. Figure 4.16 shows the simulated fitness distribution of $F(x_{\text{opt}}, \overline{x}_{\text{opt}})$. A Gaussian function was fitted to the distribution and the minimum of the approximated fitness of noise, $F_{\text{noise,min}}$, was determined using the fitted function.

As a result, inequality 4.9 reduces to

$$F(\overline{x}, \overline{x}_{\text{fit}}) \leq F(x_{\text{meas}}, \overline{x}_{\text{fit}}) - F_{\text{noise,min}}. \tag{4.11}$$

This inequality gives an upper limit for the nonideal fit $F(\overline{x}, \overline{x}_{\text{fit}})$ when the effect of noise is approximated to be as minimal as possible. Recall here that $F(x_{\text{meas}}, \overline{x}_{\text{fit}})$ is obtained from a fitting process but the minimum value for the fitness, which depends on noise, is unknown. When the effect of noise is minimized, the rest of fitness can be considered to be a consequence of a nonideal fit. This can be a case when a search algorithm has converged to a local optimum in the neighbourhood of a global optimum or the fitting procedure has met some criteria for termination. To find a maximum of, *i.e.*, the upper limit for $F(\overline{x}, \overline{x}_{\text{fit}})$, the error determination task reduces to

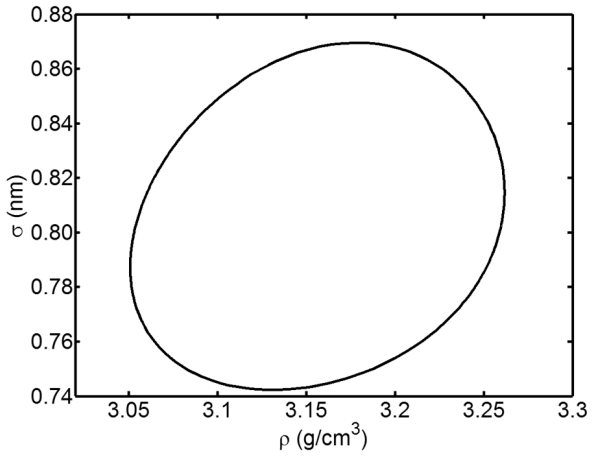
$$F(\overline{x}(p), \overline{x}_{\text{fit}}) = F(x_{\text{meas}}, \overline{x}_{\text{fit}}) - F_{\text{noise,min}}, \tag{4.12}$$

where the noiseless XRR curve $\overline{x}(p)$ is varied as the function of a parameter set p. The parameter set p was optimized radially in a selected plane until equation 4.12 was satisfied. Note that this implementation was similar to the method presented in publication IV producing ellipses in the selected planes, two of them shown in figures 4.17(a) and 4.17(b). Note that the ellipse in figure 4.17(a) is nonparaxial. Similar nonparaxiality was also observed in publication IV.

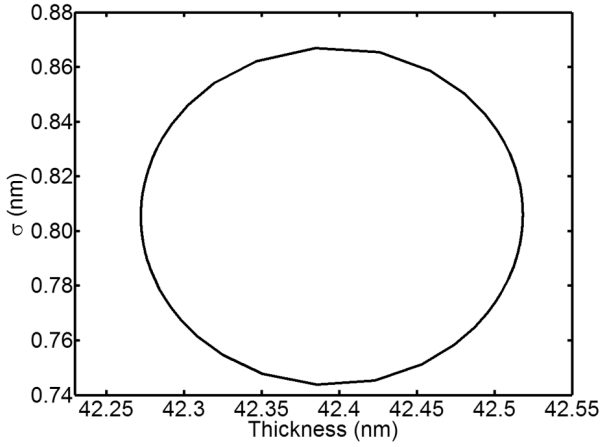
The determined parameters were 42.4±0.12 nm (±0.3%) for the thickness, 3.15±0.11 g cm^{-3} (±3.5%) for the mass density and 0.8 ± 0.06 nm (±7.5%) for the surface roughness, as shown in table 4.1. It is worth noticing now that the relative upper error limits are in the same scale with the relative lower error limits obtained in publication IV.

To compare the errors, XRR curve fitting was also performed using the χ^2 measure since χ^2 is very sensitive to the mass density. Table 4.1 shows the determined XRR parameters for χ^2 and F, the differences in the parameters and the computed error for F. As it was seen from figure 3.6(a), χ^2 cannot match the right tail of the curve and thus the difference in the thickness values is considerably higher than the calculated error. The computed error for the roughness is greater than the difference in the roughness values which supports the validity of the computed error. The difference in the mass density values, on the other hand, is slightly higher than the estimated error. Recall here that the error analysis assumed sufficiently large number of datapoints and photon counts. Therefore the inaccuracy of results given by the novel method is increased since these conditions are not met accurately. However, the applied formalism still gives reasonable results matching with the differences in the determined XRR parameters.

The presented error limits are upper limits in a case of a sufficient number of sampling points and photon counts. In practice, the determined error is an asymptotic estimate for the error (see appendix of publication V) and requires fine tuning to be done in future. The significance of the presented method is its capability to take into account a systematic error caused by a nonideal fit which may be difficult with classical methods. The novel method requires two parameters to be defined:



(a)



(b)

Figure 4.17: (a) Error limits of the solution for the mass density ρ and surface roughness σ. (b) Error limits of the solution for the thickness t and surface roughness σ [publ. V].

- The statistical significance level α. The significance level is a userdefined parameter and it can be selected to meet the required confidence level.
- The number of simulations for the distribution of the fitness of noise. The number of simulations affects the accuracy of the results and should be as large as possible.

Since the presented novel method does not contain arbitrary parameters, it gives objectively an estimate for the upper limit of the error when the convergence of a search algorithm to the optimal fit is not guaranteed.

Table 4.1: XRR parameters determined with F and χ^2 denoted as † and ‡, respectively. * denotes a parameter which was fixed during the fitting procedure. Differences are the parameter deviations in the principal Al_2O_3 layers. Error† is calculated using F measure. The parameters are here presented with three decimal accuracy for the thickness t, the AlO mass density ρ and the surface roughness σ [publ. V].

Layer	t (nm)	ρ (g cm^{-3})	σ (nm)
$Al_2O_3^\dagger$	42.393	3.152	0.804
Subtrate†	∞	2.33*	0.000
$Al_2O_3^\ddagger$	42.152	3.044	0.850
Subtrate‡	∞	2.330*	0.000
	Δt (nm)	$\Delta\rho$ (g cm^{-3})	$\Delta\sigma$ (nm)
Difference	0.242	0.108	0.046
Error†	0.123	0.105	0.064

5 Summary

Recently emerged technology for thin film fabrication, called atomic layer deposition (ALD), has reached wide attention in various fields of science and industry. This is due to the unique nature of ALD where self–limited growth of a film provides digital control of the thickness in subnanometer scale. This extremely good thickness control, on the other hand, is difficult to be verified in the given scale of the accuracy.

In the verification, x–ray reflectivity (XRR) has become a more widely accepted thin film characterization method due to its better accuracy compared to the conventional noncontact methods, such as ellipsometry. The major benefit of XRR is the short wavelength of x–rays allowing very accurate thin film thickness determination. However, XRR has suffered from the lack of dedicated and efficient analytical tools thus limiting the utilization of the full potential provided by XRR. This has been, in fact, a significant limitation in every day material research.

In XRR, material parameters are solved from a measurement by a fitting of a theoretical curve to the measurement. Important part here is also error analysis of the found parameters. This means that the scientist should be an experienced experimentalist and a theorist simultaneously if proper analytical tools are not available. However, scientist are typically concentrated either on experiments or on theoretical inspections and thus this significant multidisciplinary topic between the research areas has not gained wider attention before.

This dissertation works with this rarely studied area of x–ray reflectivity. Novel fitting algorithms and tools for error analysis were developed during this work. In the development of novel fitting algorithms, the layer models to be optimized were mimicking realistic atomic layer deposited AlO/ZnO nanolaminates. Layer models mimicking ALD AlO layers were used in the error analysis.

In the studies of the novel fitting algorithms, the rotation of coordinates was utilized during the crossover phase to reduce interparameter dependencies, called genetic linkage. The rotation of coordinates was based on Hessian covariance matrices where the eigenvectors of these matrices were forming a new basis. The crossover was performed in the rotated coordinates and the new combinations were transformed back to the original coordinates. It was shown that these novel genetic algorithms have better convergence properties than genetic algorithms without coordinate ro-

tation. It was also observed that the statistical covariance approach is clearly better than the Hessian method. The success of the statistical approach in genetic linkage reduction was extended to be investigated with the genetic algorithm using independent component analysis (ICAGA). ICAGA had better convergence properties than the classical genetic algorithm and the covariance approach. The result suggested that ICAGA is the best method to analyze multiperiodic layer structures.

The error analysis was started by studying the sensitivity properties of XRR. The work was based on empirical models describing the features of an ALD AlO structure on silicon. The sensitivity analysis was performed using fitness landscapes which were describing the behaviour of fitness as a function of the applied parameters. The analysis based on the fitness landscapes revealed that the error in the mass density parameter can compensate the error in surface roughness in a certain way. The analysis showed objectively that these parameters have some crosscorrelation in the determination as thought by many skilled experimentalists. However, the analysis was restricted to the studied case and cannot be easily generalized to other cases.

The error analysis was approached by studying the effect of Poisson noise on the accuracy of XRR analysis. A statistical p–value test was used to study a hypothesis which argued that another than the exact solution gives the best fitness. The exact solution was assumed equal to the fit given by a fitting algorithm and thus the accuracy is limited by Poisson noise. In the studied case the accuracy was found to be ± 0.09 nm with the 99% confidence level which is the lower limit, or the minimum, of the error.

The previous analysis assumed, however, that the fitting procedure is carried out perfectly and all the uncertainty in the results is a consequence of noise. This is, however, a nonvalid assumption in general. Typically the fitting procedure is carried out using a formalism and a layer model which are not fully describing instrumental and structural properties accurately. These factors together with a nonideal fit increase the fitness being more than a minimum fitness caused by Poisson noise. In this case one can get an approximation for the minimum fitness from the artificially generated fitness distribution by a separate simulation. This idea was utilized in the last publication where the original fitness obtained from the fitting was shown to be a sum of the minimum fitness and a nonideal fit. This nonideal fit was assumed to include all the nonidealities of instrumentation and the fitting thus meaning an upper limit of the error done in the analysis. Atomic layer deposited aluminium oxide layer was measured with XRR and the determined error of thickness was ± 0.12 nm with the confidence level of 99%.

The work presented in this dissertation consisted of novel genetic fitting algorithms and statistical error analysis methods. The novel genetic algorithms had considerably better robustness than the classical genetic algorithm thus significantly improving the XRR analysis of complex layer structures. Previously rarely examined error analysis in XRR was studied with simple layer structures and the methods

developed in this thesis can be used in practical every day material research. For future work the behaviour of the genetic algorithm using independent component analysis should be explored further to improve the understanding of the algorithm. Also the possibility to reduce the dimensionality of the problems using advanced methods should be investigated in order to decrease the time used in the separation of components. Also the nonlinear–fitness–space structure adaptation used in the first publication, should be developed further. Finally, finite sampling and photon counting effects have yet to be included into the error analysis. Also the implementation used in the error analysis needs improving.

References

[1] M. Leskelä and M. Ritala, Thin Solid Films **409**, 138 (2002).

[2] R. L. Puurunen, J. Appl. Phys. **97**, 121301 (2005).

[3] M. Ritala, M. Leskelä, J.-P. Dekker, C. Mutsaer, P. J. Soininen and J. Skarp, Chem. Vap. Deposition **5**, 7 (1999).

[4] P. D. Agnello, IBM J. Res. & Dev. **46**, 317 (2002).

[5] S. C. Song, Z. Zhang, C. Huffman, S. H. Bae, J. H. Sim, P. Kirsch, P. Majhi, N. Moumen and B. H. Lee, Thin Solid Films **504**, 170 (2006).

[6] E. P. Gusev, V. Narayanan and M. M. Frank, IBM J. Res. & Dev. **50**, 387 (2006).

[7] J. Aarik, A. Aidla, A.-A. Kiisler, T. Uustare and V. Sammelselg, Thin Solid Films **340**, 110 (1999).

[8] A. Yamada, B. Sang and M. Konagai, Appl. Surf. Sci. **112**, 216 (1997).

[9] J. Kim, H. Hong, S. Ghosh, K.-Y. Oh and C. Lee, Jpn. J. Appl. Phys. **42**, 1375 (2003).

[10] K.-E. Elers, V. Saanila, P. J. Soininen, W.-M. Li, J. T. Kostamo, S. Haukka, J. Juhanoja and W. F. A. Besling, Chem. Vap. Deposition **8**, 149 (2002).

[11] M. D. Groner, S. M. George, R. S. McLean and P. F. Carcia, Appl. Phys. Lett. **88**, 051907 (2006).

[12] P. F. Carcia, R. S. McLean, M. H. Reilly, M. D. Groner and S. M. George, Appl. Phys. Lett. **89**, 031915 (2006).

[13] H. Zhang and R. Solanki, J. Electrochem. Soc. **148**, F63 (2001).

[14] M. Leskelä and M. Ritala, Angew. Chem. Int. Ed. **42**, 5548 (2003).

[15] M. Leskelä, M. Kemell, K. Kukli, V. Pore, E. Santala, M. Ritala and J. Lu, Mater. Sci. Eng. **27**, 1504 (2007).

[16] D. Tonova, M. Depas and J. Vanhellemont, Thin Solid Films **288**, 64 (1996).

[17] Y. Wang and E. A. Irene, J. Vac. Sci. Technol. B **18**, 279 (2000).

[18] H. M. Cho, Y. W. Lee, I. W. Lee, D. W. Moon, H. J. Lee, B. Y. Kim, H. J. Kim, S. Y. Kim and Y. J. Cho, J. Vac. Sci. Technol. B **19**, 1144 (2001).

[19] D. Franta and I. Ohlídal, Surf. Interface Anal. **30**, 574 (2000).

[20] H. G. Tompkins, T. Zhu and E. Chen, J. Vac. Sci. Technol. A **16**, 1297 (1998).

[21] E. Chason and T. M. Mayer, Crit. Rev. Solid State Mater. Sci. **22**, 1 (1997).

[22] A. van der Lee, Solid State Sciences **2**, 257 (2000).

[23] F. Bertin, A. Chabli, E. Chiariglione, M. Burdin, M. Berger, T. Boudet, O. Lartigue and G. Ravel, Thin Solid Films **313–314**, 68 (1998).

[24] E. Bellandi, A. Elbaz, S. Spadoni, R. Piagge, C. Coccorese, G. Pavia, S. Ferrari, S. Banerjee and M. Alessandri, Thin Solid Films **450**, 120 (2004).

[25] S. Ferrari, M. Modreanu, G. Scarel and M. Fanciulli, Thin Solid Films **450**, 124 (2004).

[26] S. Kohli, C. D. Rithner, P. K. Dorhout, A. M. Dummer and C. S. Menoni, Rev. Sci. Instrum. **76**, 023906 (2005).

[27] A. van der Lee, Eur. Phys. J. B **13**, 755 (2000).

[28] M. V. Klibanov and P. E. Sacks, J. Math. Phys. **33**, 3813 (1992).

[29] W. L. Clinton, Phys. Rev. B **48**, 1 (1993).

[30] M. K. Sanyal, S. K. Sinha, A. Gibaud, K. G. Huang, B. L. Carvalho, M. Rafailovich, J. Sokolov, X. Zhao and W. Zhao, Europhys. Lett. **21**, 691 (1993).

[31] E. M. Lee, J. E. Milnes and I. C. Wong, Physica B **221**, 159 (1996).

[32] A. Asmussen and H. Riegler, J. Chem. Phys. **104**, 8159 (1996).

[33] H. J. Voorma, E. Louis, N. B. Koster, F. Bijkerk and E. Spiller, J. Appl. Phys. **81**, 6112 (1997).

[34] F. Rieutord, A. Braslau, R. Simon, H. J. Lauter and V. Pasyuk, Physica B **221**, 538 (1996).

[35] X.-L. Zhou and S.-H. Chen, Phys. Rev. E **47**, 3174 (1993).

[36] C.-H. Chou, Chinese J. Phys. **34**, 1163 (1996).

[37] C.-H. Chou, Chinese J. Phys. **34**, 58 (1996).

[38] C.-H. Chou, Physica B **233**, 130 (1997).

[39] C.-H. Chou, M. J. Regan, P. S. Pershan and X.-L. Zhou, Phys. Rev. E **55**, 7212 (1997).

[40] C.-H. Chou, Physica B **253**, 320 (1998).

[41] C.-H. Chou, Chinese J. Phys. **37**, 631 (1999).

[42] C.-H. Chou, Chinese J. Phys. **38**, 182 (2000).

[43] E. Smigiel and A. Cornet, J. Phys. D: Appl. Phys. **33**, 1757 (2000).

[44] I. R. Prudnikov, R. J. Matyi and R. D. Deslattes, J. Appl. Phys. **90**, 3338 (2001).

[45] O. Starykov and K. Sakurai, Appl. Surf. Sci. **244**, 235 (2005).

[46] P. Mikulík and T. Baumbach, Physica B **248**, 381 (1998).

[47] I. I. Samoilenko, B. M. Shchedrin and L. A. Feigin, Physica B **221**, 542 (1996).

[48] M. Geoghegan, R. A. L. Jones, D. S. Sivia, J. Penfold and A. S. Clough, Phys. Rev. E **53**, 825 (1996).

[49] D. S. Sivia and J. R. P. Webster, Physica B **248**, 327 (1998).

[50] N. F. Berk and C. F. Majkrzak, Phys. Rev. B **51**, 11296 (1995).

[51] C. F. Laub and T. L. Kuhl, J. Chem. Phys. **125**, 244702 (2006).

[52] M. K. Sanyal, J. K. Basu, A. Datta and S. Banerjee, Europhys. Lett. **36**, 265 (1996).

[53] M. R. Lovell and R. M. Richardson, Curr. Op. Coll. Interf. Sci. **4**, 197 (1999).

[54] L. G. Parratt, Phys. Rev. **95**, 359 (1954).

[55] K. Stoev and K. Sakurai, The Rigaku J. **14**, 22 (1997).

[56] A. Ulyanenkov and S. Sobolewski, J. Phys. D: Appl. Phys. **38**, A235 (2005).

[57] K. Kunz, J. Reiter, A. Götzelmann and M. Stamm, Macromolecules **26**, 4316 (1993).

[58] B. B. Luokkala, S. Garoff and R. M. Suter, Phys. Rev. E **62**, 2405 (2000).

[59] V.-O. de Haan and G. G. Drijkoningen, Phys. B Condens. Matter. **198**, 24 (1994).

[60] A. D. Dane, A. Veldhuis, D. K. G. de Boer, A. J. G. Leenaers and L. M. C. Buydens, Physica B **253**, 254 (1998).

[61] M. Wormington, C. Panaccione, K. M. Matney and D. K. Bowen, Phil. Trans. R. Soc. Lond. A **357**, 2827 (1999).

[62] A. Ulyanenkov, K. Omote and J. Harada, Physica B **283**, 237 (2000).

[63] D. Šimek, D. Rafaja and J. Kub, Mater. Struct. **8**, 16 (2001).

[64] M. Wormington, C. Panaccione, K. M. Matney and D. K. Bowen, US Patent 6192103 (2001).

[65] A. Ulyanenkov and S. Sobolewski, European Patent EP1522958 (2007).

[66] A. Ulyanenkov and S. Sobolewski, European Patent EP1522959 (2005).

[67] A. Ulyanenkov and S. Sobolewski, US Patent 7154993 (2006).

[68] A. Ulyanenkov and S. Sobolewski, US Patent 7110492 (2006).

[69] P. S. Pershan, Phys. Rev. E **50**, 2369 (1994).

[70] R. L. Puurunen, Chem. Vap. Deposition **9**, 249 (2003).

[71] K.-E. Elers, T. Blomberg, M. Peussa, B. Aitchison, S. Haukka and S. Marcus, Chem. Vap. Deposition **12**, 13 (2006).

[72] R. L. Puurunen, Chem. Vap. Deposition **9**, 327 (2003).

[73] A. W. Ott and R. P. H. Chang, Mater. Chem. Phys. **58**, 132 (1999).

[74] J. W. Elam, D. Routkevitch and S. M. George, J. Electrochem. Soc. **150**, G339 (2003).

[75] J. W. Elam, Z. A. Sechrist and S. M. George, Thin Solid Films **414**, 43 (2002).

[76] J. W. Elam and S. M. George, Chem. Mater. **15**, 1020 (2003).

[77] J. M. Jensen, A. B. Oelkers, R. Toivola and D. C. Johnson, Chem. Mater. **14**, 2276 (2002).

[78] Y. Travaly, J. Schuhmacher, A. M. Hoyas, M. Van Hove, K. Maex, T. Abell, V. Sutcliffe and A. M. Jonas, J. Appl. Phys. **97**, 084316 (2005).

[79] Z. A. Sechrist, F. H. Fabrequette, O. Heintz, T. M. Phung, D. C. Johnson and S. M. George, Chem. Mater. **17**, 3475 (2005).

[80] A. M. Hoyas, Y. Travaly, J. Schuhmacher, T. Sajavaara, C. M. Whelan, B. Eyckens, O. Richard, S. Giangrandi, B. Brijs, W. Vandervorst, K. Maex, J. P. Celis, A. M. Jonas and A. Vantomme, J. Appl. Phys. **99**, 063515 (2006).

[81] L. Zhong, W. L. Daniel, Z. Zhang, S. A. Campbell and W. L. Gladfelter, Chem. Vap. Deposition **12**, 143 (2006).

[82] F. H. Fabrequette and S. M. George, Thin Solid Films **515**, 7177 (2007).

[83] C. N. Ginestra, R. Sreenivasan, A. Karthikeyan, S. Ramanathan and P. C. McIntyre, Electrochem. Solid–State Lett. **10**, B161 (2007).

[84] K. Kukli, M. Ritala, M. Leskelä, J. Sundqvist, L. Oberbeck, J. Heitmann, U. Schröder, J. Aarik and A. Aidla, Thin Solid Films **515**, 6447 (2007).

[85] P. Bergese, E. Bontempi and L. E. Depero, Appl. Surf. Sc. **253**, 28 (2006).

[86] L. Nevot and P. Croce, Rev. Phys. Appl. **15**, 761 (1980).

[87] D. W. Marquardt, J. Soc. Industr. Appl. Math. **11**, 431 (1963).

[88] A. Corana, M. Marchesi, C. Martini and S. Ridella, ACM Trans. Math. Softw. **13**, 262 (1987).

[89] D. Mitra, F. Romeo and A. Sangiovanni-Vincentelli, Adv. Appl. Prob. **18**, 747 (1986).

[90] C. J. Bélisle, J. Appl. Prob. **29**, 885 (1992).

[91] L. Ingber, Mathl. Comput. Modelling **18**, 29 (1993).

[92] K. Gallagher and M. Sambridge, Computers & Geosci. **20**, 1229 (1994).

[93] Y.-p. Chen and D. E. Goldberg, Evol. Comput. **13**, 279 (2005).

[94] D. Wyatt and H. Lipson, GECCO 2003, LNCS **2724**, 1518 (2003).

[95] D. Whitley, M. Lunacek and J. Knight, GECCO 2004, LNCS **3103**, 294 (2004).

[96] A. Hyvärinen and E. Oja, Neural Networks **13**, 411 (2000).

[97] P. Tichaský, Z. Koldovský and E. Oja, IEEE Trans. Sign. Proc. **54**, 1189 (2006).

[98] P. Tichavský, Z. Koldovsky and E. Oja, in *IEEE/SP 13th Workshop on Statistical Signal Processing, 2005* (2005), p. 1084–1089.

[99] J. Särelä and R. Vigário, J. Mach. Learn. Res. **4**, 1447 (2003).

[100] S. Bermejo, IEEE Sign. Proc. Lett. **12**, 481 (2005).

[101] Q. Z. Zhang, N. M. Allinson and H. Yin (IEEE Symp. Comb. Evol. Comput., San Antonio TX 11–13 May, 2000), p. 33–36.

[102] M. Takahashi and H. Kita (Evol. Comput., Seoul, South Korea 27–30 May, 2001), vol. 1, p. 643–649.

[103] G. Li, K. H. Lee and K. S. Leung, *Parallel Problem Solving from Nature — PPSN IX* (Springer Berlin / Heidelberg, 2006), vol. 4193 of *Lecture Notes in Computer Science*, p. 172–182.

[104] F. J. J. Massey, J. Amer. Statist. Assoc. **46**, 68 (1951).

[105] I. P. McCarthy, Intern. J. Oper. Prod. Manag. **24**, 124 (2004).

[106] J. T. Alander, L. A. Zinchenko and S. N. Sorokin, IEEE Trans. Antennas Propag. **52**, 2932 (2004).

ISBN 978-951-22-9540-1
ISBN 978-951-22-9541-8 (PDF)
ISSN 1795-2239
ISSN 1795-4584 (PDF)